Vol. 44 APRIL, 1953 No. 4

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
JUN 2 1953
PERIODICAL READING ROOM

desk

CONTENTS

The Relation among Vocabulary, Structural Analysis, and Reading 193
JACOB TATE HUNT

Teacher Growth in Attitudes toward Behavior Problems of Children 203
MANFRED H. SCHRUPP AND CLAYTON M. GJERDE

Some Relationships between Non-intellectual Characteristics and Academic Achievement 215
JOHN P. McQUARY

Qualitative Differences in the Vocabulary Choices of Children as Revealed in a Multiple-Choice Test 229
LORRAINE P. KRUGLOV

The Validity of the Goodenough Draw-a-Man Test in Greece 244
I. TH. PAPAVASSILIOU

Book Reviews 249

Published monthly except June to September

$7.00 a year in the United States, Pan American Countries and the Philippines; Canada, $7.20; other countries, $7.40. Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

* Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLZINGER
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

INFORMATION FOR CONTRIBUTORS

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Stephen M. Corey, Teachers College, Columbia University, New York 27, N. Y.

THE JOURNAL has set regulations regarding content and style of material published, and these should be observed in the preparation of manuscripts to be submitted.

Tables and graphs.—Authors are not required to bear part of the additional cost resulting from the use of tables, formulas, and

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 44 April, 1953 Number 4

THE RELATION AMONG VOCABULARY, STRUCTURAL ANALYSIS, AND READING[1]

JACOB TATE HUNT

School of Education
University of North Carolina

The principal purpose of the investigation reported here was to discover, by the method of correlation, the relationship among vocabulary, structural analysis, and reading at the college level. By structural analysis is meant the analysis of words into their structural elements, such as prefix, stem or root word, suffix, or syllables, as an aid toward obtaining meaning.

Although few research studies have been concerned with word-elements and structural analysis, it is believed that studying word structure is an important method of developing vocabulary.(*3,4,12,13,15*) Limitations of the use of a knowledge of word-elements are pointed out by Harris(*12*) and others.(*3,17,19*) Barnes(*2*) and Buswell(*5*) report discouraging results from their short-term experiments in attempting to increase vocabulary through word-elements study. For university sophomores Carroll(*6*) obtained correlations between a morpheme (word-elements) test and other measures as follows: vocabulary, .427; intelligence, .251; and Latin participation, .489.

Artley(*1*) lists ten types of contextual aids, one of which is word-elements. Gray(*10*) points out that contextual clues may practically insure identification or, on the other hand, merely limit the possibilities. In a discussion of improvement of vocabulary, Dickinson(*7*) states that although adults learn by guessing from context, children do not. On the day after sixty-seven college freshmen had read an eleven-page essay on the "Luxury of Integrity," Sachs(*16*) found that only thirteen were able to

[1] This article is based upon a portion of the writer's Ph.D. dissertation on file in the library of The University of California, Berkeley. The study was made under the guidance of Dr. Luther C. Gilbert.

write a correct definition of the term 'integrity,' which had been used eleven times. Dilla(*8*) states that few college freshmen make even moderately sound definitions of such words as 'civility' and 'compromise' and that reading the context gives little help. Correlations of .50 between context and intelligence and of .56 between context and vocabulary are reported by Gibbons.(*9*) Summaries of the relationship among intelligence, vocabulary, and reading indicate that correlations between any two of these are typically between .40 and .70.(*11,14,18*)

PROCEDURE

This investigation was conducted in the Spring semester of 1948, with students enrolled in the introductory course in Educational Psychology at the University of California as the subjects. Administration of an intelligence test and occasionally a reading test is part of the usual class activity for the course. Tests specially constructed for the study were tried out in experimental form with other students before revision and use with the subjects of the investigation. All tests were administered and scored by the writer.

Since it was desired to construct two tests using some words whose meanings were unknown to the subjects, it was necessary to administer vocabulary tests to locate unknown words. For this purpose, the vocabulary section of the reading test and a specially constructed eighty-word, five-choice vocabulary test were used. Preliminary tests with undergraduate and graduate students enrolled in other courses had yielded these eighty difficult and relatively unknown words. The special vocabulary test was utilized only to obtain words for two other tests, context and word-meaning construction. Only those words missed by eighty per cent or more were selected for these two tests. Since there were five choices for each item in the two vocabulary tests, it was assumed that a word missed at a ratio of four to one would be virtually unknown. No item in either of the vocabulary tests was omitted by more than five per cent of the students participating.

THE TESTS EMPLOYED

Standardized tests used in the study were the *A.C.E. Psychological Examination*, 1941 series, and the *Coöperative Reading*

Comprehension Test, C2, Upper Level, Form T. The reading test gives scores for vocabulary, speed of comprehension, and level of comprehension. Tests constructed by the writer consisted of vocabulary, rate and comprehension, and the four tests employed to infer ability in structural analysis: word-derivation, word-elements, word-meaning construction, and context. The reliability coefficient of each test or subtest is .60 or higher, with a median r of .82.

The word-meaning construction test consists of twenty words, with each of which is supplied the various etymological elements and their meanings. (Sample item: tergiversant—*tergum*, back; *vertere*, to turn; *ant*, agent or doer). Using this supplied information, the testee is to construct an acceptable definition of the word's meaning.

The test of context consists of supplying the meaning for an unknown word embedded in meaningful context of about fifty words or more. The items are so constructed that more than one approach or technique may be used for obtaining the meaning of each word. In certain cases it is expected that a knowledge of word-elements will be helpful. A sample item follows:

> One of the children's favorite stories was that of the huge ants which had controlled a certain little continent for centuries. These ants were friendly enough so long as plant food was plentiful, but during plant scarcities they built ingenious man-traps disguised as pool halls, bookie joints, etc. The hapless victim was then carried off to one of the hot springs for boiling. These *anthropophagous* ants were civilized in their own way and hunted only from necessity rather than for pleasure.

The word-elements test has sixty-six items, consisting of twenty-two four-choice items in each of three divisions of the test, prefixes, suffixes, and combining forms. The material for the rate-comprehension test is of a popular science type not requiring special background. Rate is obtained by dividing the number of words (2273) by the time required to read the article; comprehension is determined by the responses to sixteen four-choice questions.

The word-derivation test consists of twenty common words selected because their elements have common meanings and had been used in a straight-forward manner etymologically. A check on the student's knowledge of the words was made by a multiple-

choice test. A correct free response to the item 'auditorium' to measure word-derivation requires the identification of the stem or root (*audio, audi, audit*) with its meaning of 'to hear' and the suffix, *orium*, meaning 'a place for.' Items were not considered correct unless both the various elements and accepted meanings for all elements were given.

THE POPULATION

Among the one hundred sixty-eight students participating were sixty-nine men and ninety-nine women, classified as follows: sophomores, ten; juniors, one hundred thirteen; and seniors, forty-five. Graduate and special students were excluded from the present study. The mean and standard deviation on the A.C.E. were 124.74 and 19.11. Combining samples available from 1941, 1946, and 1947 sections of the course from which the population was obtained yielded a mean of 125.61 and a standard deviation of 21.98 for three hundred fifty-seven students on this same series. A score of 125 is at the seventy-fifth percentile for college freshmen. Combining percentile ranks by college classification into one distribution resulted in median percentile ranks for the Coöperative Reading Test as follows: vocabulary, 49.5; level of comprehension, 58.4; and speed of comprehension, 66.

Thirty-eight different major subjects or areas were represented, indicating that the group was made up of students with heterogeneous interests and academic backgrounds. Major areas reported most frequently were General Curriculum, English, Physical Education, Psychology, History, and Mathematics. Of the seventy-nine who had studied Latin for one year or more, thirty-two were men and forty-seven were women. All had studied at least one modern foreign language. There is no reason to believe that this is not a reasonably good cross-section of the upper division students of the University.

THE RESULTS

Coefficients of correlation were computed for fifteen variables by means of scattergrams and the formula for r when deviations are taken from the assumed means of the two distributions. Because a large number had not studied Latin, correlations between Latin study and the other measures were computed by

the bi-serial r formula. Zero-order coefficients for thirteen variables are given in the top right half of Table 1; partial r's with intelligence held constant are given in the lower left half of the same table. Correlations of Latin and modern language with the other variables were found to be typically low or insignificant and have been omitted from the tables in order to conserve space. A summary of the relationships for these two is given at the end of this section. Although our chief interest is in the relationship of various measures of reading and vocabulary with the specially constructed tests of structural analysis, more extensive relationships were established for purposes of comparison. For 166 degrees of freedom, an r of .199 is significant at the one per cent level.

First-order correlations, upper right, Table 1, reveal that although correlations are not high, intelligence, vocabulary, reading, and measures of structural analysis are all related.

TABLE 1.—ZERO-ORDER AND PARTIAL COEFFICIENTS OF CORRELATION FOR 13 VARIABLES (N = 168)*

		Zero-order Correlations											
	Partial Correlations (Intelligence Constant)	2	3	4	5	6	7	8	9	10	11	12	13
1. ACE		56	48	57	26	33	45	39	30	47	42	41	34
2. Vocab.			58	58	26	29	44	35	34	39	53	30	38
3. Level		43		71	19	39	31	28	25	27	50	30	34
4. Speed		38	61		38	38	37	29	30	35	38	39	37
5. Rate		14	08	30		−03	23	20	20	18	13	20	09
6. Comp.		14	28	25	−12		28	23	10	30	25	18	25
7. World-Ele.		24	11	16	13	15		87	81	88	44	36	52
8. Prefix		17	12	09	12	12	85		44	55	31	27	41
9. Suffix		22	13	16	14	01	79	37		43	21	30	40
10. CF		17	05	24	06	18	85	45	34		44	34	41
11. Context		39	37	19	03	13	31	17	10	30		42	51
12. W-M Const.		10	12	20	11	05	21	14	21	18	30		46
13. Word-Deriv.		24	28	19	07	16	43	33	34	31	43	37	

* Decimal points and plus signs omitted.

Read zero-order correlations from the top right half of the table; partial correlations with intelligence constant, lower left half.

Correlations among intelligence, vocabulary, and reading are about as expected from previous research. The short tests of reading rate and comprehension do not show relationships with other measures to so high a degree as do speed and level of comprehension.

Marked relationships were found between the total word-elements test and word-derivation (.52), intelligence (.45), vocabulary (.44), and context (.44). Word-elements are related at the one per cent level with all other variables except for number of courses of modern language studied. The marked correlation with vocabulary and context and the higher relationship with the latter than with other reading measures suggest the value of knowledge of word-elements as a means of increasing word knowledge. In general, the subtest of combining forms is found to be somewhat more closely related to other measures than either prefixes or suffixes.

Context shows a marked relationship (.40 or above) with seven other measures. Highest correlations with context are for vocabulary and word-derivation. The substantial correlations with these two, with word-elements, and with word-meaning construction suggest that structural analysis is an important means of obtaining word meanings. Combining forms, however, are as closely related to context as is the total word-elements test, indicating that those elements more nearly resembling words and having more exact meanings are most usable for constructing meanings of words.

Word-meaning construction is most closely related to word-derivation, context, and intelligence. Those who can recognize the greatest number of elements of familiar words are superior in constructing the meanings of unknown words even when the meanings of the parts of the words are supplied. Next in magnitude are the relationships to speed of comprehension and to word-elements with its three subtests. Vocabulary, level of comprehension, and modern language show correlations of approximately .30 with word-meaning construction. Although some few other correlations are significant they are quite low.

Word-derivation is significantly related to all other measures except rate. Its highest correlations are with word-elements and context. The relatively high correlation with word-elements is expected. Somewhat surprising is the correlation of the same

size with context, showing a higher relationship than with vocabulary or intelligence. Of considerable interest is the relationship of word-derivation with Latin (.33) and modern language study (.40).

Although most of the zero-order r's are not very high, it is understood that these are as high as they are because of common factors. Partial correlation represents the net relationship between two measures when an element common to both is partialled out and no longer affects either test. Partial r's with intelligence held constant are shown in the lower left half of Table 1; with vocabulary constant, upper right half of Table 2; with word-elements constant, lower left half of Table 2.

TABLE 2.—COEFFICIENTS OF CORRELATION WITH VOCABULARY AND WORD-ELEMENTS PARTIALLED OUT (N = 168)*

		Vocabulary Constant												
	Word-Elements Constant	1	2	3	4	5	6	7	8	9	10	11	12	13
1. ACE				24	36	14	22	28	25	14	34	18	31	17
2. Vocab.		44												
3. Level		42	52		57	05	29	07	10	07	06	28	16	16
4. Speed		48	50	66		29	28	16	12	13	16	11	27	20
5. Rate		15	19	13	33		−13	13	13	13	08	−01	13	−01
6. Comp.		23	19	34	31	−10		16	14	01	22	12	10	16
7. Word-Ele.									85	78	86	27	26	42
8. Prefix										37	48	16	19	32
9. Suffix											34	03	22	32
10. CF												30	25	31
11. Context.		27	42	42	26	03	15						29	39
12. W-M const.		30	17	21	30	13	09					31		39
13. Word-Deriv.		13	19	22	22	−03	13					36	34	

* Decimal points and plus signs omitted.

Read correlations with vocabulary constant from the top right half of the table; with word-elements constant, lower left half.

The general effect of partialling out intelligence, vocabulary, or word-elements is to lower most of the correlations and to lower a considerable number below the level of significance. Partialling out either intelligence or vocabulary has a greater effect on the correlations than does partialling out word-elements except in certain cases such as in the correlation between word-derivation

and Latin when a knowledge of word-elements is an important factor in each. Holding either intelligence or vocabulary constant lowers considerably the correlations between word-elements and the other measures.

Correlations of word-elements with context, word-meaning construction, word-derivation, and Latin remain significant at the one per cent level when either intelligence or vocabulary is held constant. The relationship of word-elements with vocabulary is reduced from .44 to .26 when intelligence is held constant and the correlation of .45 between word-elements and intelligence is reduced to .28 when vocabulary is held constant. The word-elements test is not significantly related to either power or speed of reading when either intelligence or vocabulary is controlled.

When either intelligence or vocabulary is held constant, context is significantly related to level of comprehension, word-elements, combining forms, word-meaning construction, word-derivation, and Latin. Context and vocabulary are related when intelligence is held constant, but context is not related to intelligence when vocabulary is held constant. Holding word-elements constant causes correlations of context with comprehension and Latin to fall below the significance level.

Controlling intelligence leaves word-meaning construction significantly related to speed of reading, word-elements, suffixes, context, word-derivation, and modern language. Similar relationships, plus those with intelligence and combining forms, exist when vocabulary is held constant.

Word-derivation continues to be related to the other measures of structural analysis and to both Latin and modern language study when either intelligence, vocabulary, or word-elements is held constant. Holding word-elements constant typically had a greater effect on the correlations between word-derivation and the other measures, especially those of structural analysis and language study, than did intelligence or vocabulary.

Latin and modern language study are not related to vocabulary or reading except indirectly through word analysis. Low but significant relationships exist between Latin study and rate (.21), word-elements (.27), context (.23), and word-derivation (.33). Although a significant relationship exists between Latin and these four measures when either intelligence or vocabulary is held constant, only word-derivation remains significant when

word-elements are partialled out. Modern language study is related to only word-meaning construction (.29) and word-derivation (.40). A significant relationship remains between modern language study and these two measures when any one of the three variables is held constant.

SUMMARY AND CONCLUSIONS

The aim of this study was to determine the relationships among vocabulary, structural analysis, and reading. Ability in structural analysis was inferred from tests of word-elements, context, word-derivation, and word-meaning construction. Zero-order and partial r's were computed for one hundred sixty-eight university students on fifteen variables.

The following conclusions appear valid in light of the findings and limitations of this study: (1) Vocabulary, structural analysis, and reading show moderate interrelationships. Structural analysis is related somewhat less to vocabulary and reading than are vocabulary and reading related to each other. (2) Of the four tests of structural ability, context is most closely related to vocabulary and reading. (3) The tests of structural ability are interrelated and tend to have a cumulative effect. (4) The more intelligent students are likely to possess greater ability in using structural analysis than the less intelligent, even among a somewhat select university group. Ability to use structural analysis is more than a matter of general intellectual ability, since it tends to be related to reading and vocabulary even when the effect of intelligence is controlled. (5) Level of comprehension tends to be related more to the analytical skills involved in structural analysis than does speed of comprehension. (6) Latin and modern language study have a low or negligible relationship to structural analysis, reading, or vocabulary. (7) Teaching designed to improve the student's methods of work attack should be encouraged.

BIBLIOGRAPHY

1) A. Sterl Artley. "Word meanings through context," *Elementary English Review*, 20: 68–75, 1943.

2) Melvin W. Barnes. "Root words and current meanings," *School and Society*, 56: 610–11, 1942.

3) G. L. and Eva Bond. *Developmental Reading in High School.* New York: Macmillan Co., 1941.

4) Leona M. Burfield. "Basic Facts and Techniques Essential in Dealing with Word-perception Difficulties in High Schools and Junior Colleges," *Appraisal of Current Practices*, Supplementary Educational Monographs, No. 61. Chicago: University of Chicago Press, 1945.

5) Guy T. Buswell. *Remedial Reading at the College and Adult Levels*, Supplementary Educational Monographs, No. 50. Chicago: University of Chicago Press, 1939.

6) J. B. Carroll. "Knowledge of English roots and affixes as related to vocabulary and Latin study," *Journal of Educational Research*, 34: 102–11, 1940.

7) D. Dickinson. "Importance of vocabulary in reading," *Elementary School Journal*, 20: 539, 1920.

8) Geraldine P. Dilla. "Words, words, words," *School and Society*, 51: 243–47, 1940.

9) Helen D. Gibbons. "The ability of college freshmen to construct the meaning of a strange word from the context in which it appears," *Journal of Experimental Education*, 9: 29–35, 1940.

10) William S. Gray. *On Their Own in Reading*. Chicago: Scott, Foresman & Co., 1948.

11) ———. "Reading," *Encyclopedia of Educational Research* (Monroe, Editor). New York: Macmillan, 1949.

12) A. J. Harris. *How to Increase Reading Ability*. New York: Longmans, Green & Co., 1947.

13) Nelle F. Kirchner. "Reading in the secondary school," *English Journal*, 34: 329–33, 1945.

14) C. M. Louttit. *Clinical Psychology*. New York: Harper and Bros., 1936.

15) Joseph Mersand. *Do Our Pupils Know What We Are Talking About?* New York: Modern Education Chapbooks No. 1, 1939.

16) H. J. Sachs. "Reading method of acquiring vocabulary," *Journal of Educational Research*, 36: 457–464, 1943.

17) Ruth Strang. *Problems in the Improvement of Reading*. Lancaster, Pa.: Science Press, 1938.

18) ———. "Relationship between certain aspects of intelligence and certain aspects of reading," *Educational and Psychological Measurement*, 3: 355–359, 1943.

19) C. J. Ruger. "The effect of first-year Latin upon knowledge of English words of Latin derivation," *School and Society*, 18: 260–70, 1923.

TEACHER GROWTH IN ATTITUDES TOWARD BEHAVIOR PROBLEMS OF CHILDREN

MANFRED H. SCHRUPP and CLAYTON M. GJERDE

San Diego State College

Wickman's study of teachers' attitudes toward behavior problems of children, published in 1928,[1] has been quoted ever since as evidence that teachers are unable to recognize as serious those problems so designated by clinicians. The very fact that his study has so frequently been quoted should have been influential in changing teachers' attitudes. Certainly teacher education programs have devoted increased time and effort to the development of a 'mental hygiene viewpoint.' It would therefore seem desirable to determine whether the attitudes of teachers of today correlate as poorly with those of mental hygienists as the professional literature implies. The major purpose of this study was to compare present-day teachers' attitudes with attitudes of earlier teachers and mental hygienists and with mental hygienists of today. To facilitate direct comparisons with Wickman's results, his procedures were repeated as closely as practicable.

In his study, Wickman had a group of five hundred eleven teachers and a group of thirty mental hygienists rate the seriousness of fifty behavior traits of children. He found that the ratings made by the teachers correlated about zero with the ratings made by the mental hygienists, and that there appeared to be wide discrepancies between the two groups with regard to the kinds of behavior problems which were considered serious. In interpreting his study, it should be noted that the ratings were made by the teachers and hygienists on the basis of distinctly different instructions. Wickman was careful to point this out in his report:

"The techniques employed for measuring the reactions of the mental hygienists to behavior disorders, therefore, differed in certain respects from the methods employed in measuring teachers' reactions. In the rating scales for teachers the adoption of three precautionary techniques for controlling the teachers' responses will be recalled . . . : (1) the directions to teachers for

[1] E. K. Wickman, *Children's Behavior and Teachers' Attitudes*, New York: Commonwealth Fund. 1928.

rating were phrased in such a way as to secure responses to the *present* problem, and the question of the significance of the present behavior disorder upon the *future* development of the child, though possibly unavoidably implied, was not definitely raised. The task set was to rate the degree of maladjustment represented by the immediate problem. (2) Care was also taken to establish in the teachers a mental set for responding to the "seriousness of" the amount of "difficulty produced by" the particular type of troublesome behavior. The assumption was that the degree to which teachers found a certain problem serious, difficult, or undesirable represented the amount of attention they directed to the problem and the effort exerted towards its modification. (3) Then, too, in order to elicit the first, unrationalized reactions, the teachers were instructed to rate as rapidly as possible, and a time limit was imposed for completing the ratings.

"The precautionary techniques utilized in measuring the teachers' attitudes were exactly reversed in eliciting the attitudes of the mental hygienists. (1) Instead of evaluating the *present* problem, the mental hygienists were directed to rate the significance of the problem in terms of its effect on the *future* life of the child . . . (2) Though the terms "seriousness" and "difficulty" of a problem were retained in the directions for rating, the concept of the "importance" of the behavior problems was emphasized and replaced the concepts of "consequence" and "undesirableness." (3) Instead of issuing instructions for rating as rapidly as possible and imposing a time limit for completing the ratings, as administering the scale to teachers, an attempt was made to elicit from the mental hygienists responses that were intellectually controlled and evaluated. The directions read, 'Try to make this a professional opionion that is as free as possible from your emotional reactions.' "[2]

The substance of this quotation has often been overlooked when the results of Wickman's research have been discussed.[3] Ellis

---

[2] Wickman, op cit., pp. 119–121.

[3] It is interesting to note that modern authors continue to quote Wickman's study without indicating the fact that the ratings made by teachers and clinicians are not directly comparable because of major differences in directions to the two groups for rating the seriousness of the problems listed. Texts in psychology and educational psychology, published in the

and Miller[4] attempted to correct this difficulty in their 1936 study with Denver teachers by using Wickman's schedule, but modifying the directions so that they were essentially the same as those presented by Wickman to the mental hygienists. They found a correlation of .49 between the Denver teachers and Wickman's mental hygienists. Unfortunately it was not possible to determine whether this larger correlation between teachers and hygienists represented a change in point of view on the part of this teacher group, or whether it resulted from the change in directions. Mitchell,[5] in a study conducted in 1940 with teachers from the same school systems used by Wickman, employed a modification of the original Wickman scale, as well as a modification of the directions. Sixty-three mental hygienists also rated the traits, following the same directions used with the teachers. He reported a correlation of .70 between teachers and mental hygienists. He also reported a correlation of .21 between Wickman's teachers of 1927 and his mental hygienists of 1940, indicating a possible shift in the opinions of mental hygienists. Here, again, the higher relationship between teachers and mental hygienists could be due to the fact that these two groups followed the same directions in making their ratings.

In the present study, Wickman's schedules B-4 and B-5[6] were employed with the teachers and mental hygienists, respectively, without modification of any kind, either in the schedule of behavior traits itself or in the directions. It was therefore possible to make direct comparisons of present findings with those of twenty-four years ago. It should be remembered, however, that in this study, as in Wickman's, the results do not permit a direct comparison of teachers and clinicians as professional

last decade, were pulled from the library shelves until a dozen were found which cited Wickman's study. Of these, seven failed to mention this important fact, while stressing the differences in attitudes between teachers and clinicians. Three mentioned a difference in 'point of view' of the two groups but failed to indicate the differences in directions. Only two gave a clear statement of the differing directions used by Wickman.

[4] D. B. Ellis, and L. W. Miller, "Teachers' Attitudes and Behavior Problems," *Journal of Educational Psychology*, 27:501–511, Oct., 1936.

[5] John C. Mitchell, "A Study of Teachers' and Mental Hygienists' Ratings of Certain Behavior Problems of Children," *Journal of Educational Research*, 36:292–307, December, 1942.

[6] Wickman, op. cit., pp. 205–211.

groups. Here, as in Wickman's study, the careful and considered judgment of the clinicians was used as the most nearly ideal criterion possible for evaluating the teachers' rapid, unrationalized reactions to the behavior problems.

Schedule B-4 and a personal data sheet were sent to one hundred ninety-nine teachers, selected at random from regularly employed teachers in the secondary and elementary schools of San Diego, California, through the coöperation of the Department of Research of the city schools. Of the one hundred ninety-nine schedules distributed, one hundred twenty-seven (63.8 per cent) were returned, and 119 (59.8 per cent) were usable. Of the one hundred nineteen teachers responding, fifty-nine were from the elementary level, and sixty from the secondary level.

Thirty-seven mental hygienists completed schedule B-5, and they were employed by public school guidance agencies or clinics, as were those in the Wickman study. Thirty-one of them were from the San Diego City School Guidance Bureau, while six were employed as clinicians by the Long Beach (California) City Schools. The group was composed of one psychiatrist, twelve psychologists, twenty-four school social workers and visiting teachers. As will be shown later, the ratings by these clinicians correlated highly (.80 to .88) with clinicians in 1927 and in 1940, which suggests that these groups were quite comparable in their attitudes toward behavior problems of children, insofar as the Wickman scale measured them.

As in the Wickman study, the respondents were asked to rate the behavior traits by marking on a line, scaled on the teachers' form from 'Of no consequence' to 'An extremely grave problem,' and on the clinicians' form from 'Of no importance at all' to 'Of extremely great importance.' The responses were quantified by the use of a scale ranging from 0 to 20, as described by Wickman,[7] thus permitting the determination of mean scale scores on each trait for teachers and clinicians. It was possible then to compute product-moment correlations between the means of the ratings by the teacher group and the means of ratings by Wickman's clinicians, Mitchell's clinicians, and those involved in the present study. In addition, similar correlations were determined between the means of ratings made by groups of teachers in 1927,

[7] Wickman, op. cit., p. 97.

1940, and 1951, and between the means of the ratings made by clinician groups in the same years. These correlations are presented in Table I.

TABLE I.—CLINICIAN AND TEACHER INTERCORRELATIONS ON ATTITUDES TOWARD BEHAVIOR PROBLEMS OF CHILDREN

	1951 Tchs.	1940 Tchs.	1927 Tchs.	1951 Clin.	1940 Clin.
1927 Clin.	.43**	.35*	−.04	.88**	.80**
1940 Clin.	.61**	.70**	.21	.88**	
1951 Clin.	.56**	.54**	.09		
1927 Tchs.	.76**	.78**			
1940 Tchs.	.81**				

* Significant at the five per cent level.
** Significant at the one per cent level.

Using the ratings of the clinicians of 1927 as a criterion, there is evidence of definite increase in agreement between clinician and teacher attitudes, as the correlation of .43 is statistically significant at the one per cent level. Although the greatest change appears to be between 1927 and 1940, this cannot be assumed since Mitchell's 1940 group of teachers rated the behavior problems following the same directions employed in the case of the clinician group. For the same reason, the correlation between the 1940 clinicians' ratings and the 1940 teachers' ratings (.70) may not be directly comparable to those obtained in 1927 and 1951. However, the 1951 teachers' ratings correlated .61 with those of the 1940 clinicians, a considerable and significant gain over the .21 reported by Mitchell between the 1927 teachers' ratings and the 1940 clinicians' ratings.

If the ratings of the clinicians involved in the present study (hereafter referred to as '1951 clinicians') are used as a criterion, we again see a significant increase in the correlations between teachers' and clinicians' ratings. The 1927 teachers' ratings correlated only .09 with those of 1951 clinicians, while the ratings by 1951 teachers correlated .56 (significant at the one per cent level) with those of the 1951 clinician group. Here, too, the increase appears at first glance to be greater from 1927 to 1940 than from 1940 to 1951, but this may again be due to the fact that the 1940 teachers followed essentially the same directions in

rating the traits as did the 1951 clinicians. This was not true in the case of the 1951 teachers.

Correlations between ratings by the three clinician groups (.88, .88, .80) are all significant at the one per cent level, and of a high order. For the teacher groups the same is true, with correlations ranging from .76 to .81. As mentioned earlier, this suggests that the teacher and clinician groups are distinct groups rather than being random samples from the same population. To determine if this was true for the 1951 teacher and clinician groups, each group was randomly divided into two subgroups of approximately equal size. Product-moment correlations were then computed between the mean scale scores of the subgroups. This correlation was .94 for the two teacher subgroups, and also .94 for the two clinician subgroups, again indicating that teachers and clinicians were distinct and fairly homogeneous groups.

In general, the correlations presented in Table I suggest that a definite increase in agreement between teachers and clinicians took place between 1927 and 1951, and that this increased agreement is probably due primarily to a change in the attitudes of teachers rather than of clinicians. Assuming (1) that the Wickman scale actually measures attitudes toward behavior problems of children, (2) that clinicians' ratings constitute a valid criterion, and (3) that the teachers studied are reasonably representative groups, it can be concluded that teachers' attitudes of this kind have improved significantly in the last twenty-four years.

Another method of indicating the relationships between two groups is by comparison of ranks. Wickman's study is frequently cited as evidence of lack of teacher concern for children's behavior traits which are indicative of tendencies toward shyness and withdrawal. This conclusion has been based in part upon a comparison of those of the fifty behavior traits which were rated among the ten most serious by one group and the ten least serious by the other group. Table II summarizes these data for the three teacher and three clinician groups of 1927, 1940, and 1951.

In 1927, there were five traits in all which were rated among the ten most serious by one group and among the ten least serious by the other. In 1951, there were no such extreme disagreements. Two traits, 'disobedience' and 'destroying,' which were rated among the ten least serious by the 1927 clinicians were rated

TABLE II.—BEHAVIOR TRAITS RANKED AMONG THE TEN MOST SERIOUS BY ONE GROUP AND AMONG THE TEN LEAST SERIOUS BY THE OTHER GROUP

	1927 Teachers	1940 Teachers	1951 Teachers
1927 Clin.	Masturbation*	Destroying*	Disobedience*
	Destroying*	Masturbation*	Destroying*
	Unsocialness	Sensitiveness	
	Overcritical		
	Sensitive		
1940 Clin.	Unsocialness	(none)	Disobedience*
	Overcritical		
1951 Clin.	Masturbation*	Masturbation*	(none)
	Unsocialness	Shyness	
	Overcritical		
	Shyness		

* Rated among ten most serious by teacher group, and ten least serious by clinician group. Not starred—vice versa.

among the ten most serious by 1951 teachers. As has been pointed out by Ellis and Miller,[8] this may be in part justified when one considers the nature of the work of the two groups and the directions which the two groups were asked to follow in their ratings. Perhaps the 1951 clinicians recognized this, since they ranked 'destroying' 15, and 'disobedience' 24.

Table III gives, for each of the fifty traits, the means of the ratings by the 1951 teachers and clinicians (columns 1 and 2). These means have been ranked (columns 3 and 4), and the differences in ranks determined (column 5). Similar information for 1927 teachers and clinicians, as reported by Wickman, are also indicated (columns 6, 7, and 8).

Examination of the rank orders indicates that some significant disagreements still exist between teachers and clinicians in their rating of seriousness of behavior traits. Among the fifty traits, there were sixteen for which the rank difference between ratings by 1951 teachers and clinicians was 15 or greater. These are listed in Table IV in the order of their rank differences.

The evidence in Table IV suggests that teachers, compared with clinicians, still tend to be more concerned with those behavior traits which appear to be transgressions against orderliness and,

[8] Ellis and Miller, op. cit., 508.

TABLE III.—MEANS AND RANKS OF MEANS OF RATINGS OF FIFTY BEHAVIOR PROBLEMS OF CHILDREN BY CLINICIAN AND TEACHER GROUPS

Behavior Problems	(1) Means 1951 Clin.	(2) Means 1951 Teach.	(3) Ranks 1951 Clin.	(4) Ranks 1951 Teach.	(5) Rank Difference	(6) Ranks 1927 Clin.	(7) Ranks 1927 Teach.	(8) Rank Difference
Unsocial, withdrawing	17.1	12.9	1	12.5	−11.5	1	40.5	−39.5
Unhappy, depressed	16.6	14.2	2	5	− 3	3	22.5	−19.5
Cruelty, bullying	15.2	15.5	3	1	2	6	8	− 2
Fearfulness	13.9	11.6	4	26	−22	5	36	−31
Suspiciousness	13.2	10.6	5	32.5	−27.5	2	37	−35
Suggestible	12.9	12.2	6	18.5	−12.5	8	28	−20
Nervousness	12.1	11.8	7	23	−16	18.5	20.5	− 2
Easily discouraged	12.0	12.2	8	18.5	−10.5	7	22.5	−15.5
Shyness	11.9	9.8	9	40	−31	13.5	50	−36.5
Overcritical of others	11.3	11.0	10	29.5	−19.5	9	45	−36
Temper tantrums	11.2	12.7	11	15	− 4	17	13	4
Stealing	11.2	15.2	12	2	10	13.5	2	11.5
Dreaminess	11.2	9.9	13	38.5	−25.5	18.5	40.5	−22
Resentfulness	11.1	13.1	14	11	3	4	29	−25
Destroying school materials	11.1	14.0	15	6	9	45	10	35
Truancy	10.9	12.9	16.5	12.5	4	22	6	16
Sensitiveness	10.9	10.0	16.5	37	−20.5	10	48	−38
Enuresis	10.8	11.6	18	26	− 8	27	19	8
Domineering, overbearing	10.2	11.6	19.5	26	− 6.5	11	32.5	−21.5
Unreliable, irresponsible	10.2	13.5	19.5	7	12.5	21	12	9
Physical coward	10.1	10.8	21	31	−10	15	31	−18
Sullenness, sulkiness	10.0	11.8	22	23	− 1	12	35	−23

Lack of interest in work	9.9	11.2	23	28	− 5	26	14	12
Untruthfulness	9.8	13.4	24	8	16	23	5	18
Cheating	9.6	12.4	25	16	9	23	9	14
Heterosexual activity	9.2	14.5	26	3	23	25	1	24
Selfishness, unsportsmanly	9.0	11.9	27	21	6	16	24.5	− 8.5
Laziness	8.5	11.0	28	29.5	− 1.5	35.5	16.5	19
Inattention	8.4	10.3	29	35	− 6	34	26	8
Impertinence, defiance	8.1	14.3	30.5	4	26.5	37.5	7	30.5
Quarrelsomeness	8.1	12.3	30.5	17	13.5	31	27	4
Imaginative lying	8.1	7.9	32	48	−16	33	42	− 9
Disobedience	7.7	13.5	33	9	24	41	11	30
Obscene notes, pictures, etc.	7.3	13.2	34	10	24	28.5	4	24.5
Carelessness in work	7.2	10.5	35	34	1	37.5	24.5	13
Slovenly in appearance	7.2	9.2	36	42	− 6	35.5	34	1.5
Tardiness	7.0	8.5	37	47	−10	43	30	13
Thoughtlessness	6.9	8.8	38	44	− 6	39	38	1
Restlessness	6.8	8.6	39	46	− 7	41	49	− 8
Impudence, rudeness	6.5	12.8	40	14	26	32	16.5	15.5
Stubbornness, contrariness	6.4	10.2	41.5	36	5.5	20	32.5	−12.5
Silliness	6.4	9.9	41.5	38.5	3	30	39	− 9
Masturbation	6.3	11.8	43	23	20	41	3	38
Disorderliness	5.8	12.0	44	20	24	46	20.5	25.5
Tattling	5.6	8.6	45	45	0	28.5	46.5	−18
Profanity	4.9	10.6	46	32.5	13.5	47	15	32
Smoking	4.9	9.4	47	41	6	49	18	31
Inquisitiveness	4.2	7.8	48	49	− 1	44	43.5	.5
Interrupting, talkativeness	3.8	9.1	49	43	6	48	43.5	4.5
Whispering	2.4	7.5	50	50	0	50	46.5	3.5

UNIVERSITY OF MICHIGAN LIBRARY

TABLE IV.—TRAITS ON WHICH GREATEST DISAGREEMENT APPEARS WHEN RATED BY 1951 TEACHERS AND CLINICIANS

Traits Rated More Serious by Teachers	Rank Difference	Traits Rated More Serious by Clinicians	Rank Difference
1. Impertinence, defiance	26.5	1. Shyness	31
2. Impudence, rudeness	26	2. Suspiciousness	27.5
3. Obscene notes, pictures, etc.	24	3. Dreaminess	25.5
4. Disobedience	24	4. Fearfulness	22
5. Disorderliness	24	5. Sensitiveness	20.5
6. Heterosexual activity	23	6. Overcritical of others	19
7. Masturbation	20	7. Imaginative lying	16
8. Untruthfulness	16	8. Nervousness	16

perhaps, morality, and less concerned with those traits which appear to be related to withdrawal. Again, it must be remembered that this difference may be due, in part at least, to the differences in directions given to the two groups. However, it seems likely that the differences reflect a real difference in attitude between the two groups, since Mitchell's[9] data also support this conclusion in spite of the fact that his groups followed identical directions in making the ratings.

The difference between 1951 teachers and clinicians, as indicated in Table IV, does not appear as large as the difference between 1927 teachers and clinicians as shown in Table V. It will be noted that, for the 1927 groups, there were twenty-eight traits in which there was a rank difference of 15 or more, as compared with sixteen for the 1951 groups. The magnitude of the difference between teachers' and clinicians' ratings also tended to be greater for the 1927 groups.

SUMMARY

This study was an attempt to re-examine certain conclusions, presented by Wickman in 1928 and still widely cited, as to the attitude of teachers towards the behavior problems of children. In an effort to closely approximate a replication of the Wickman study, the experimental design duplicated that used by Wickman. Therefore, it was possible to make direct comparisons between

[9] Mitchell, op. cit.

TABLE V.—TRAITS ON WHICH GREATEST DISAGREEMENT APPEARED WHEN RATED BY 1927 TEACHERS AND CLINICIANS

Traits Rated More Serious by Teachers	Rank Difference	Traits Rated More Serious by Clinicians	Rank Difference
1. Masturbation	38	1. Unsocial, withdrawing	39.5
2. Destroying	35	2. Sensitiveness	38
3. Profanity	32	3. Shyness	36.5
4. Smoking	31	4. Overcritical of others	36
5. Impertinence, defiance	30.5	5. Suspiciousness	35
6. Disobedience	30	6. Fearfulness	31
7. Disorderliness	25.5	7. Resentfulness	25
8. Obscene notes, pictures, etc.	24.5	8. Sullenness, sulkiness	23
9. Heterosexual activity	24	9. Domineering, overbearing	21.5
10. Laziness	19	10. Dreaminess	21.5
11. Untruthfulness	18	11. Suggestible	20
12. Truancy	16	12. Unhappy, depressed	19.5
13. Impudence	15.5	13. Tattling	17
		14. Physical coward	16
		15. Easily discouraged	15.5

the results of the two studies, in spite of their common limitations. In other words it was appropriate to compare ratings by teachers of 1927 with those by teachers of 1951 by examining their relationships with the criterion ratings, even though clinicians and teachers as professional groups should not be compared directly.

This comparison showed, insofar as the groups studied can be considered representative, that the attitudes of teachers of 1951 agreed much more closely with the 'ideal criterion' than did teachers of 1927. This was evidenced primarily in three ways:

1) The correlation between the means of ratings of the 1951 teacher and clinician groups was .56, as compared with −.04 in Wickman's study of 1928.

2) None of the traits listed among the ten most serious by one group was listed among the ten least serious by the other group. Wickman found five traits so rated.

3) As shown in Tables IV and V the extent of disagreement between teachers and clinicians in 1951 was not as great as was true in 1927.

Although clinician and teacher groups agreed much more closely in 1951 than in 1927, definite disagreements were still evident. While it is difficult to determine the degree to which the disagreements were a result of experimental design, it is interesting to note that the direction of the disagreement was similar to that cited by Wickman. Teachers, when compared with clinicians still appeared to be less concerned about behavior traits associated with withdrawal and more concerned about those which appear to be transgressions against orderliness and, perhaps, morality.

This study has demonstrated the value of repeating earlier studies with no significant modification of experimental design, especially when those earlier studies frequently serve as the basis for generalizations. The chief difference between the present study and that of Wickman was with respect to the geographical location of the populations studied. This variation, however, should not materially affect the validity of the following general conclusions:

1) The attitudes of 1951 teachers toward the behavior problems of children were much more in agreement with the criterion attitudes established by clinicians than was true for 1927 teachers. This fact should be considered carefully by authors of textbooks in psychology and educational psychology. In fairness to both professional groups, the data from this study, as well as Wickman's study, should be interpreted in the light of the experimental design. It is likely that teachers' attitudes will never approximate very closely the 'ideal criterion attitudes' as established by clinicians, since good teachers will always need to be concerned about temporary, but disturbing, behavior in the classroom.

2) Disagreements between attitudes of teachers and the criterion attitude established by clinicians, though not as pronounced as in 1927, still exist, and these disagreements are of the same nature as those pointed out by Wickman. Those responsible for teacher education, both pre-service and in-service, evidently need to continue to emphasize what might be called 'a mental-hygiene viewpoint.'

SOME RELATIONSHIPS BETWEEN NON-INTELLECTUAL CHARACTERISTICS AND ACADEMIC ACHIEVEMENT

JOHN P. McQUARY

University of Wisconsin

The purpose of this investigation was to determine the factor pattern underlying variables assumed to be related to scholastic achievement of male freshmen at the University of Wisconsin. Many of the twenty-three variables are so-called non-intellectual characteristics, such as size of home community, occupational level of the father, educational level of the parents, and number of siblings. Their relation to each other and to academic achievement (grade points earned) was determined, utilizing techniques of correlation and factor analysis.

Two hypotheses were tested:

1) Achievement in college is significantly related to certain non-intellectual variables.

2) The non-intellectual variables can be grouped into several factors.

THE SAMPLE

The sample used was one hundred seventy-four first semester freshmen males at the University of Wisconsin who had sought the services of the Student Counseling Center (either for general counseling or reading and study help) in their first semester. The services of this organization are available to any student and are on a voluntary basis. The clients of the University of Wisconsin Student Counseling Center are given consecutive case numbers after they have had an initial preliminary interview. The method used for selection in this study was to examine each of these case folders after September 15, 1948. All those men who were first semester freshmen at the University of Wisconsin in the first semester of the 1948–49 school year and those who were first semester freshmen in the first semester of the 1949–50 school year and for whom were available all the data to be used as variables were accepted. Therefore, the sample was determined by completeness of data.

MEASURES USED

The descriptions of the twenty-three variables follow.

Variable Number	Description
1	The quantitative reasoning raw score, American Council on Education Psychological Examination, (*ACE Q*), (1947 edition)
2	The linguistic raw score, (*ACE L*)
3	*Speed of reading* scaled score, Coöperative English Test (Form Q, 1940 edition)
4	*Level of Comprehension* scaled score, Coöperative English Test
5	*Vocabulary* scaled score, Coöperative English Test
6	*High-school Percentile Rank*
7	*Occupational level of the father.* This rating was made according to the job classification system used in the *Dictionary of Occupational Titles.* The highest rating corresponded to a highly professional occupation, while the lowest indicated an unskilled job.
8	*Size of community* in which the individual had spent the majority of his life. The highest score corresponded to a city of over 100,000 population. The lowest was assigned to those who had spent most of their lives on farms.
9	*Educational level of the father.* The score taken for this variable was the number of years of formal schooling which the father of the student had had.
10	*Educational level of the mother.* Scores were obtained in the same manner as for Variable 9.
11	*Number of siblings.* A score of 1 meant that the individual was an only child, 2 meant that he was one of two children in the family, 3 meant that he was one of three children, etc.
12	*Position among the siblings.* A score of 1 was assigned to those individuals who were the first child in the family, a score of 2 to those who were the second child, a score of 3 to those who were the third, etc.
13	*High-school extra-curricular participation.* A score on this variable was taken as the number of seven activity areas (literary, dramatics, debating, music and art, athletics, student government, and others) in which each individual had some participation during high school. The extent or importance of the participation was not accounted for.
14	*Number of illnesses and diseases.* This variable indicated the total number of illnesses or diseases which the individual

had had or currently had. A health check list included not only the usual childhood diseases but also such things as use or need for glasses, hernia, backaches, headaches, epilepsy, etc. Those health difficulties not listed could be added.

15 *Hours studied per week.* The score on this was the student's estimate of the average number of hours per week he studied, excluding class hours.

16 *Type of financial support.* A score was obtained from a seven-point rating scale where one extreme was the totally self-supporting student, and the other, the student who was completely supported by his parents.

17 *Number of credits carried.* This variable indicated the total number of credits carried by a student throughout the first semester. It did not take into account those credits which might have been dropped prior to the deadline for dropping courses without failure, i.e., prior to the end of the eighth week of the semester.

18 *Grade points earned.* The score on this variable was obtained in the following manner: 3 grade points were given for each credit of A work, 2 grade points for each credit of B work, 1 for each credit of C, 0 grade points for each credit of D, $-\frac{1}{2}$ grade points for each credit of E, and -1 for each credit of F work.

19 *Social introversion-extroversion.* A high score is in the direction of introversion, a low score in the direction of extroversion as measured by the social introversion-extroversion scale of the Minnesota Multiphasic Personality Inventory.

20 *Mother's age at son's birth.*

21 *Vocational certainty.* A rating was made according to four categories: very certain of choice, somewhat certain of choice, somewhat uncertain of choice, and very uncertain of choice.

22 *Foreign-born parent.* This variable had only two possible responses: first, at least one of the student's parents was born outside of the United States, second, neither parent was foreign born.

23 *Veteran or non-veteran.*

PROCEDURE

The distributions of the twenty-three variables were first normalized by converting the original scores to stanines. The pairs of twenty-one of these variables were then used to obtain Pearson product-moment correlations. The other two variables, numbers 22 and 23, were paired with each of the others to calcu-

late bi-serial correlations; these two were then paired to obtain a tetrachoric correlation. These intercorrelations made up the entries of the original correlation matrix.[1] The multiple group method of factoring was chosen as the factoring technique. Communalities were estimated to be equal to the highest correlation in each column. The factoring operation was according to the procedure outline by Harris and Schmid (*1*).

The factoring was stopped after seven factors had been extracted. The criterion that the standard deviation of the residuals should equal or be less than the standard error of a zero correlation for the sample was used as the basis for stopping the factoring process. This criterion was met after seven factors had been removed.

The factor matrix obtained from this procedure was turned into an orthogonal solution (F). Single-plane and radial rotations were made on columns of the F matrix until an oblique solution (V) was obtained. (See Table I).[2] According to Thurstone's criteria for simple structure, the V matrix is a 'neat' solution. Note that from 11–18 variables fall within the range of $\pm .10$ on each factor.

INTERPRETATION OF THE FACTORS

Factor C will be interpreted first since it represents familiar relationships between intellectual characteristics and academic achievement.

Factor C.—	
ACE Q	.28
ACE L	.38
Speed of Reading	.42
Level of Comprehension	.34
Vocabulary	.39
High-school Percentile Rank	.50
Credits carried	.34
Grade points earned	.56

[1] The author's doctoral dissertation, *Relationships between Non-intellectual Characteristics and Academic Achievement,* is deposited in the University of Wisconsin Graduate Library.

[2] The original correlation matrix, the residual matrix, the orthogonal solution (F), and the matrix of direction cosines ($\wedge$) are deposited with the American Documentation Institute, Washington, D. C.

This factor might be called 'Academic Performance Predictor.' Every 'intellectual' variable used in this study appears on this

TABLE I

	A	B	C	D	E	F	G
1	07	22	28	11	07	14	09
2	−06	24	38	−15	−03	−03	27
3	00	07	42	−03	−04	07	36
4	01	02	34	−03	−10	08	47
5	−05	05	39	−06	−01	−04	40
6	02	−01	50	51	09	00	−01
7	−16	−09	−09	09	08	20	30
8	04	00	02	49	−05	00	−08
9	02	39	06	08	11	−10	04
10	10	48	06	−02	03	09	−01
11	09	00	−09	11	−03	22	50
12	07	02	−06	−12	03	09	52
13	−04	09	06	66	60	03	06
14	23	04	09	−10	03	−03	−31
15	−09	−11	04	02	06	−31	−09
16	−09	−08	−02	−14	09	03	08
17	−04	04	34	06	07	09	−08
18	05	−13	56	21	10	−07	09
19	04	−08	11	−04	−45	08	34
20	03	−01	09	−04	16	−10	14
21	21	−02	07	−07	−01	−02	−22
22	27	−53	07	−08	−29	−14	−09
23	−41	10	−01	09	−15	−13	11
Number of Variables Within ±.10	18	16	14	14	17	17	11

factor. High quantitative and linguistic scores on the ACE, high scores on speed of reading, level of comprehension, and vocabulary are linked with credits carried and grade points earned in the first semester of college. This factor is made up of the type of variables which are generally used to make predictions of academic success by means of multiple regression equations.

It is interesting to note that the variable, high-school percentile rank, has the highest factor loading on this factor. It has

the largest weight on the multiple prediction equation used at the University of Wisconsin (ACE total score is the other variable used), and is the best single predictor of college success as reported by many investigators.

Factor B.—

ACE Q	.22
ACE L	.24
Educational level of the father	.39
Educational level of the mother	.48
Foreign-born parent	−.53

This factor seems to represent *Social-Class Intelligence.* Davis and Havighurst (*2*) have reported large differences between social-class groups on various scholastic ability tests. They conclude that such tests are measuring only a narrow range of mental activities—the range of academic problems. They report that the available scholastic ability tests include numerous items which do not pertain to the culture of approximately sixty per cent of all Americans who grow up in the lower social status groups, which are heavily populated with immigrant families and those of little formal education.

There is general agreement that two major determiners of behavior may be involved in solving problems included in scholastic ability tests. These are: (1) the individual's genetic equipment (heredity) and (2) the individual's particular cultural experiences, training, and motivation (environment). This factor throws no light on the heredity-environment controversy since it may be interpreted in two ways. One may reason that since highly educated parents tend to be of high intelligence, the factor loadings of educational level of the parents, does, in part, suggest that the potentialities of intellectual capacity of the children are transmitted genetically. On the contrary, these same factor loadings may indicate an environmental influence, since the children of highly educated parents probably have a greater opportunity for practice in solving academic-type problems and probably have greater motivation to do so. The negative factor loading (−.53) implies that children having a foreign-born parent may be penalized on the traditional scholastic ability test because of the cultural setting in which they have been reared. It suggests that the effect of a language handicap

upon the test performance of children with foreign-born parents needs to be taken into account.

At first it may seem surprising that the occupational level of the father did not appear upon this factor. This may be because the occupation of the father may not indicate either the level of genetic endowment or the presence of academic motivation and training in the home situation. It more likely may be a result of selection among college students. Sandiford (*3*) found a corresponding relationship between the intelligence of children and the occupation of the father highly evident among elementary-school students, smaller in the case of high-school pupils, and negligible in the case of college students. This is due perhaps to a continual increase in selection which lops off the very low socio-economic groups.

Factor D.—

High-school Percentile Rank	.51
Size of community	.49
High-school extra-curricular participation	.66
Grade points earned	.21

Factor D appears to represent the 'Participating Urban Scholar.' This factor is one of two which has the variable, grade points earned, as a factor loading. It is interesting to note, however, that while Factor C is made up entirely of intellectual variables, this factor has non-intellectual characteristics also linked with academic achievement. High rank in high-school class, spending most of one's life in a large community, and participating in a large number of high-school extra-curricular activities are grouped with grade points earned in the first semester of college.

It has been pointed out previously that rank in high-school class is often the best single variable for the prediction of college achievement. Perhaps the loading of high-school percentile rank on this factor coupled with the loading on size of community represent a greater reliability of high-school marks in the large high schools. In addition, the high schools in the larger communities may have a wider selection of background courses, thus giving their graduates a better preparation for their academic work at the University.

With these two variables, i.e., rank in high-school class and size of community, being linked with high-school extra-curricular participation, one wonders about the underlying differences between extra-curricular participation in the large high school and the small high school. Because of a fewer number of students to choose from, perhaps, in the small high school, the opportunity for participating in school activities may be more available and may require less effort on the student's part to participate while, in the large high school, participation in extra-curricular activities may be a better indication of leadership. In addition, there is a possibility that to participate in a great number of extra-curricular activities in a large high school, a student has to be more aggressive. This willingness to fight for what he wants may carry over into his academic work in college. Finally, the adjustment which takes place when coming to a very large university may be easier for this group than for the students coming from small high schools. This greater facility to bridge the gap between high school and college may pay dividends in academic achievement in the first semester at the university.

Factor E.—

High-school extra-curricular participation	.60
Introversion-extroversion	−.45
Foreign-born parent	−.24

This factor appears to be 'social extroversion.' The two largest loadings are high-school extra-curricular participation and Drake's Social I-E scale (in the direction of extroversion). In addition, there is a small loading on non-foreign-born parents.

Psychologists have been trying to explain and measure introversion-extroversion for some time. In 1924, Freyd (*4*) listed fifty-four traits which had been used to describe introverts and extroverts up to that date. Two years later Heidbreder (*5*) selected a random sample of two hundred students from nine hundred general psychology students at the University of Minnesota who had rated themselves and two associates on Freyd's list of fifty-four characteristics of introversion and extroversion. She found that these individuals did not fall into distinct groups, introverts and extroverts, but into a single group which took the general shape of the normal probability

curve. The fifty-four characteristics were also listed in order of discriminative power. Five of the twenty-four best (the first, seventh, ninth, seventeenth, and twenty-first) were concerned with activity participation.

To make further validation of the Social I-E scale of the Minnesota Multiphasic Personality Inventory, Drake and Thiede (*6*) took high-school activity participation as a criterion. They had 594 female students at the University of Wisconsin list those high-school extra-curricular activities in which they had participated. There were six possible activity categories. (These were identical with the categories used in this study except that in this research, one additional category was added for the listing of those activities which did not fit into one of the other six.) They found highly significant differences, using Student's t test, between the mean scores on the Social I-E scale for those female students who had participated in four, five, or six of these high-school activities and the mean scores for those who had participated in none, one, or two activities.

The Guilfords (*7*) showed by a factor analysis that the items in introversion-extroversion tests were not measuring a single dimension. A typical set of thirty-six questions such as had been traditionally used to diagnose tendencies toward introversion and extroversion was administered to nine hundred thirty students. Four factors were interpreted as (1) social introversion-extroversion, (2) emotional sensitiveness, (3) impulsiveness, and (4) interest in self. Their social introversion-extroversion factor appeared to indicate that at one end of a scale an individual seeks to withdraw, to remove himself from or not permit himself social contacts and social responsibilities; at the other end, the individual seeks out social participation and depends upon it for his satisfactions. This is somewhat indicative of a continuum of amount of activity participation. Drake's Social I-E scale appears to be getting at the same type of thing.

The extroversion factors obtained from this study seems to add further validation of the Social I-E scale of the Minnesota Multiphasic Personality Inventory. It also raises the question of whether children having a foreign-born parent are penalized in some way so that activity participation is restricted. The particular culture in which the student is reared may influence his behavioral development in many ways. The operation of

environmental forces may not only vary the amount but the kind of social activities available to him. Perhaps the children of an immigrant parent do not have enough money to participate in some high-school activities, or are required to work after school. Group membership in some activities may not be available to these students. In addition, a lack of a feeling of 'belongingness' because of cultural differences or minority insecurities may prevent some of them from attempting to participate in high-school extra-curricular activities.

Factor F.—

Occupational level of the father	.20
Siblings (toward small families)	.22
Hours studied per week	−.31

This seems to be an 'Academic and Financial Security' factor. Male students who come from small, professional families tend to spend less time on their college studies. It is interesting to note that grade points earned and scholastic ability do not appear on this factor. This appears to be a 'no-need-to-strive' type of factor. Perhaps the fact that they come from professional families may lessen the drive for financial and vocational succ which students from the lower socio-economic groups so often strive for. For the latter group, striving for good academic marks is a necessary step to take to assure the realization of their financial and vocational success. Those from small professional families may have known no financial insecurity. There is the possibility that this group may more reasonably get involved in activities other than schoolwork because they have the financial means to do so.

Factor G.—

ACE L	.27
Speed of Reading	.36
Level of Compression	.47
Vocabulary	.40
Occupational level of the father	.30
Siblings	.50
Position among siblings	.52
Health	−.31
Introversion-extroversion	.34
Vocational certainty	−.22

This factor suggests that there is an 'Introvertive-reader' type of college student. It is interesting to note that grades and the quantitative aspects of scholastic ability as measured by the ACE do not appear on this factor. This reading-linguistic ability is possessed by the older children of small professional families. Linked with this reading-linguistic ability are uncertainty of vocational choice, a poor health history as evidenced by a self-rating, and social introversion.

It has been suggested that one of the chief reasons why reading is important to many people is on account of its 'instrumental' uses (*8*). A person reads because he expects to gain something from the activity. He may read to satisfy some need for a greater sense of security. This factor suggests that there is a type of college student who because of problems of vocational difficulties, perplexities about health or about popularity and social competency uses reading as an instrument to satisfy these needs or to 'escape' from them. He may read in order to improve his vocational efficiency, to apply each health remedy he reads about, or to seek answers for his inability to participate satisfactorily in social situations. Case-study literature, particularly autobiographies, often reveal similar uses of reading (*9*). Readers typically lack something which their companions possess —good health, friends, skill in athletics, etc.

One might have expected to find that this factor included a lack of financial security as well as uncertainties in the vocational, social, and health areas. The failure for this to show up as a significant loading may be accounted for by the lack of reading opportunity. Investigations have shown a definite association between occupational status of the father and the reading achievement and reading interests of the children. McNee (*10*) has reported that students whose fathers belong to the professional occupational groups show superior achievement in reading. Similarly, Beggs (*11*) has shown that children of the professional groups read more than others.

Factor A.—This factor will not be discussed in this paper. Its loadings are small and none of the variables have had extensive investigations which might offer some explanations. It would appear that this factor might be an 'Age' factor since the age of the students was not controlled in this investigation. The ϕ matrix (intercorrelations of the factors) appears in Table II.

environmental forces may not only vary the amount but the kind of social activities available to him. Perhaps the children of an immigrant parent do not have enough money to participate in some high-school activities, or are required to work after school. Group membership in some activities may not be available to these students. In addition, a lack of a feeling of 'belongingness' because of cultural differences or minority insecurities may prevent some of them from attempting to participate in high-school extra-curricular activities.

Factor F.—

Occupational level of the father	.20
Siblings (toward small families)	.22
Hours studied per week	−.31

This seems to be an 'Academic and Financial Security' factor. Male students who come from small, professional families tend to spend less time on their college studies. It is interesting to note that grade points earned and scholastic ability do not appear on this factor. This appears to be a 'no-need-to-strive' type of factor. Perhaps the fact that they come from professional families may lessen the drive for financial and vocational succ which students from the lower socio-economic groups so often strive for. For the latter group, striving for good academic marks is a necessary step to take to assure the realization of their financial and vocational success. Those from small professional families may have known no financial insecurity. There is the possibility that this group may more reasonably get involved in activities other than schoolwork because they have the financial means to do so.

Factor G.—

ACE L	.27
Speed of Reading	.36
Level of Compression	.47
Vocabulary	.40
Occupational level of the father	.30
Siblings	.50
Position among siblings	.52
Health	−.31
Introversion-extroversion	.34
Vocational certainty	−.22

This factor suggests that there is an 'Introvertive-reader' type of college student. It is interesting to note that grades and the quantitative aspects of scholastic ability as measured by the ACE do not appear on this factor. This reading-linguistic ability is possessed by the older children of small professional families. Linked with this reading-linguistic ability are uncertainty of vocational choice, a poor health history as evidenced by a self-rating, and social introversion.

It has been suggested that one of the chief reasons why reading is important to many people is on account of its 'instrumental' uses (*8*). A person reads because he expects to gain something from the activity. He may read to satisfy some need for a greater sense of security. This factor suggests that there is a type of college student who because of problems of vocational difficulties, perplexities about health or about popularity and social competency uses reading as an instrument to satisfy these needs or to 'escape' from them. He may read in order to improve his vocational efficiency, to apply each health remedy he reads about, or to seek answers for his inability to participate satisfactorily in social situations. Case-study literature, particularly autobiographies, often reveal similar uses of reading (*9*). Readers typically lack something which their companions possess —good health, friends, skill in athletics, etc.

One might have expected to find that this factor included a lack of financial security as well as uncertainties in the vocational, social, and health areas. The failure for this to show up as a significant loading may be accounted for by the lack of reading opportunity. Investigations have shown a definite association between occupational status of the father and the reading achievement and reading interests of the children. McNee (*10*) has reported that students whose fathers belong to the professional occupational groups show superior achievement in reading. Similarly, Beggs (*11*) has shown that children of the professional groups read more than others.

Factor A.—This factor will not be discussed in this paper. Its loadings are small and none of the variables have had extensive investigations which might offer some explanations. It would appear that this factor might be an 'Age' factor since the age of the students was not controlled in this investigation. The ϕ matrix (intercorrelations of the factors) appears in Table II.

TABLE II

ϕ

	A	B	C	D	E	F	G
A	1.00	.22	−.59	.53	.08	.27	.46
B	.22	1.00	.21	.00	−.07	−.43	.36
C	−.59	.21	1.00	−.44	.12	−.02	.03
D	.53	.00	−.44	1.00	−.01	.13	.14
E	.08	−.07	.12	−.01	1.00	.46	.31
F	.27	−.43	−.02	.13	.46	1.00	.09
G	.46	.36	.03	.14	.31	.09	1.00

Seven intercorrelations are zero indicating that the factors are at right angles to each other; seven are near enough to zero so that, for all practical purposes, they can be considered independent of each other, and seven have reasonably large correlations. Of the last group, three occur with Factor A which has not been discussed. If it is an 'Age' factor, the intercorrelations seem logical. One would expect the positive relationship between social-class intelligence and reading as evidenced by the .36 correlation between Factor B and Factor G. The positive .46 correlation between Factor E and Factor F can be accounted for by the fact that the extrovert would tend to participate in many activities which would limit the number of hours he could engage in studying. The −.44 correlation between Factors C and D and the −.43 correlation between Factors B and F cannot be accounted for.

SUMMARY AND CONCLUSIONS

1) This study suggests that types of data about individuals that might be of especial interest to the cultural antropologist, the sociologist, and the social psychologist may be amenable to interpretation from factor analysis. This study is particularly interesting since it combines such data with paper-and-pencil data commonly used by the educational psychologist.

2) Judged by Thurstone's criteria for simple structure, a highly satisfactory solution was secured. The factors are readily interpretable as well as clearly defined by the data. The study illustrates the possibilities of the oblique solution.

3) Two factors are necessary to account for the correlation of grade points earned with the other variables used in this study.

One factor (C) groups together those types of variables (primarily pencil-and-paper tests and rank in high-school class) that have commonly been found either singly or in combination to be relatively efficient predictors of academic success. The other (D), however, grouping high-school rank, size of community, high-school extra-curricular participation, and grades suggests a somewhat different from the usual and possibly supplementary approach to predicting academic success.

4) All the common factor variance of the ACE cannot be accounted for by one factor. The second factor (B) is an interesting linkage of psychological and cultural variables.

5) There appears to be an introvertive-reader type of college student whose reading is not directly related to academic achievement. His reading may be used for seeking answers to personal questions or difficulties.

6) The social extroversion factor appears to support the validity of the Social I.E. scale of the MMPI as an index of participation in activities.

7) Hours studied per week does not appear on either Factor C or Factor D, the two factors having grades as loadings.

8) Since the sample was not random, this needs to be considered when interpreting the data.

9) A similar study using female university students is in progress.

REFERENCES

1) C. W. Harris and John Schmid, Jr. "Further application of the principles of direct rotation in factor analysis." *Journal of Experimental Education*, 18: 175–193, March, 1950.

2) W. A. Davis and R. J. Havighurst. "The measurement of mental systems (Can intelligence be measured?)" *Scientific Monthly*, 68: 301–316, April, 1948.

3) Peter Sandiford. "Paternal occupations and intelligence of offspring." *School and Society*, 33: 117–119, January, 1926.

4) Max Freyd. "Introverts and extroverts." *Psychological Review*, 31: 74–87, January, 1924.

5) Edna Heidbreder. "Measuring introversion and extroversion." *Journal of Abnormal and Social Psychology*, 21: 120–134, July, 1926.

6) L. E. Drake and W. B. Thiede. "Further validation of the Social I. E. Scale for the Minnesota Multiphasic Personality Inventory." *Journal of Educational Research*, 41, 551–556, March, 1948.

7) J. P. Guilford and Ruth B. Guilford. "Personality factors S, E, and M and their measurement." *The Journal of Psychology*, 2: 109–127, 1936.

8) D. Waples, B. Berelson, and F. Bradshaw. *What Reading Does To People.* Chicago: University of Chicago Press, 1940.

9) C. A. Murchison. *A History of Psychology in Autobiography.* Worcester: Clark University Press, 1932.

10) Marcia A. McNee. *Factors Related to Reading Achievement of College Freshmen.* Master's Thesis, University of Chicago, 1931.

11) Berenice Beggs. "Does environment determine a child's reading?" *School and Community*, 14: 16–18, January, 1928.

QUALITATIVE DIFFERENCES IN THE VOCABULARY CHOICES OF CHILDREN AS REVEALED IN A MULTIPLE-CHOICE TEST

LORRAINE P. KRUGLOV

The City College
New York

The fact that individual vocabularies differ in quality, as well as in range, has only recently been considered in evaluating a person's level of intellectual functioning, although psychologists have been aware of qualitative differences in vocabulary responses for almost half a century. Traditionally, vocabulary differences have been considered in terms of range—how many words one can define—while the dimension of the quality of these word definitions has been largely ignored.

Feifel and Lorge (*2*), in a review of the literature, indicated that Chambers in 1904, Binet and Simon in 1916, Terman in 1916, Piaget in 1926, and others at more recent dates noted that vocabulary changes not only in range but also in quality with age. Terman recognized that children's definitions varied in quality, some children giving definitions in terms of use and appearance, others in terms of synonyms. He felt that these qualitative differences indicated the maturity of apperceptive processes; nevertheless, the scoring system of the Terman Revision of the Binet-Simon Scale failed to consider these qualitative differences.

In a recent study Feifel and Lorge (*2*) described the types of definitions characteristic of different age levels (six to fourteen) with a view toward pointing up characteristic differences in thinking for these age levels. They classified responses to the forty-five item vocabulary subtest of the Form L Stanford-Binet Test into five categories as follows: (1) Synonym, (2) Use, Description, and Use and Description, (3) Explanation, (4) Demonstration, Repetition, Illustration and Inferior Explanation, and (5) Error. The mean number of responses within each category was compared among age groups. The main findings were:

1) Older children significantly more often employ the synonym type of definition than do younger children.

2) Children six to nine significantly more often give use and description types of responses than do older children.

3) Children six to nine significantly less frequently employ explanation types of responses than do older children.

4) Demonstration, illustration, inferior explanation and repetition types of responses increase from six to ten and decrease thereafter.

In general the younger children, aged six to nine, tend to perceive words as concrete ideas—in terms of use, description, demonstration—and do not generalize; the older children, aged ten to fourteen, tend to emphasize the abstract or 'class' feature of word meanings in their use of the synonym and explanation type responses. These qualitative differences in vocabulary, reflecting the conceptual level or mode of thinking of the child, are a decided refinement on the 'range of vocabulary' which has been considered a major index of intellectual functioning to date.

Feifel and Lorge demonstrated these qualitative differences through an analysis of responses to a recall type test containing items like "What does ORANGE mean?" The question arose as to whether such qualitative differences would be revealed on a recognition type test where all the choices for each item were correct but of different conceptual levels. Studies have shown that a person's recognition vocabulary exceeds his recall vocabulary as far as vocabulary range is concerned. In the same way a child may recognize a definition at a more 'mature' level and choose it as the best definition of a word, even though he does not himself define the word at that level. The six year old might define an 'orange' as "you eat it" but he might recognize that it is "a fruit." Thus the six-year-old's recognition vocabulary might be qualitatively the same as the ten-year-old's.

If, however, the child's choices of responses are a reflection of his conceptual level, he might not recognize the more 'mature' definition as such, but might choose as the best meaning of a word a definition characteristic of his own conceptual level.

In addition to the theoretical issue, there are practical advantages to using a multiple choice test. A major drawback in the estimation of a child's conceptual level from an analysis such as Feifel and Lorge made is the need for trained scorers and the time-consuming nature of such scoring. Using a recognition type, multiple-choice test would eliminate the need for personnel trained in qualitative scoring; in fact, such a test could be machine scored.

The present study is an attempt to find out whether the qualitative differences in vocabulary response, characteristic of the different age levels, as found by Feifel and Lorge in analyzing responses to a recall type vocabulary test, would also be demonstrated in a recognition type vocabulary test. Does the child's recognition vocabulary differ qualitatively from his recall vocabulary? Does he recognize more 'mature' definitions (i.e. those at higher developmental levels), as better than definitions characteristic of his own developmental level when both types of definition are presented to him?

A multiple-choice vocabulary test was constructed in which three or four of the five choices were correct according to the traditional scoring system for the Stanford-Binet, but of different qualitative levels according to the Feifel-Lorge study. Children, at different age levels, were asked to choose the 'best' meaning for each word. Responses at different age levels were analyzed to see whether qualitative differences hold up when a recognition type test is used.

CONSTRUCTION OF THE TEST

A number of problems arose in the course of constructing the multiple-choice vocabulary test. First of all, all words do not lend themselves to Feifel's four-fold classification system. Feifel, himself, noted that not all words permit a full range of qualitative differences to appear in the verbatim responses (*1*, p. 67). This is especially so for words that are not nouns and for the more difficult nouns. In addition, in a paper-and-pencil test, demonstration type responses had to be omitted. For these reasons the kind of words to be included in the multiple-choice test was limited.

The question of the form of the choices raised a second problem. Test construction theory recommends that all the choices be parallel in form, but because of the nature of this test, it was believed that the choices should parallel the verbatim responses given in the recall situation rather than parallel each other.

A third, and the most important, problem concerned the 'difficulty' or the familiarity of the words in the test to the subjects. To satisfy the requirements of a good vocabulary test, none of the five choices should be more difficult than the stem word. Moreover, for this particular test, those responses con-

sidered to represent a higher conceptual level, i.e., synonym and explanation, were to be no more difficult in terms of vocabulary range than responses representing the lower conceptual levels, i.e., use and description, and repetition-illustration-inferior explanation, or than the error responses. The *Thorndike-Lorge Teachers Word Book of 30,000 Words* (*3*) was used to obtain the frequency of occurrence in written English of each of the words used in the test. Where a response consisted of two or more words, the lowest frequency of occurrence of any word in the response was taken as the frequency of occurrence of the entire response.

One limitation of the use of the Thorndike word frequencies is that these frequencies are based upon occurrence in written English, while the test choices are based upon spoken English, and the two are not strictly parallel in frequency. Although frequency of occurrence is not the only, and may not even be the best index of word difficulty, it was felt that setting as criteria the requirements that the stem be no more frequent in occurrence than the five choices, and that the choices indicative of higher conceptual levels be no more frequent in occurrence than choices indicative of the lower conceptual levels would tend to equate at least one component of word difficulty.

The concern with frequency of occurrence was necessary to ensure that any qualitative differences found for the different age levels were not a function of the vocabulary range of that age level, but rather a function of the thinking processes of children of that age. If, for example, the synonyms and explanations were out of the vocabulary range of a particular age group while the descriptions and uses, and repetition-illustration-inferior explanations were within the vocabulary range of that age, vocabulary range (or frequency of occurrence) rather than mode of thinking might be the crucial factor in choosing the best meaning for a word.

A ten item vocabulary test was constructed using as stem words items 1, 2, 4, 6, 7, 10, 12, 14, 17, and 23 of the Stanford-Binet Form L vocabulary test. The five choices for each item were chosen from verbatim responses made by children to the recall type item. The test is presented in Figure 1.

Altogether there were fifty choices in the ten word vocabulary test. These choices are classified in Table 1.

TABLE 1.—CLASSIFICATION OF CHOICES IN VOCABULARY TEST BY TYPE OF RESPONSE

Type of Response	Number	Per cent
Synonym	10	20
Explanation	3	6
Use and Description	14	28
Repetition, Illustration, Inferior explanation	9	18
Errors	14	28
	50	100

FIGURE 1

Name_____ Age_____ Grade_____

Below you will see a word in capitals. Five people were asked what the word means. The five choices after the word are what the five people said. You are to decide which of the answers is best. Sometimes more than one choice will tell correctly what the word means, but you should choose only the one that you think is best. Write its letter at the end of the line. The example shows how to do it.

DOG A. you drink it
B. it's a small animal
C. something wooden
D. a small house
E. it means a pet dog..... B

The best answer for DOG is a small animal, so the letter B was put in the space at the end of the line. Do the others in the same way. Go ahead and do the rest.

1. ORANGE A. it's a fruit
B. you eat it
C. it's round and yellow
D. orange-like in orange juice
E. a kind of monkey....._____

2. ENVELOPE A. it's a letter envelope
B. a wild animal
C. a container for a letter
D. white folded paper
E. something to mail things in....._____

3. PUDDLE A. muddy water
B. like a small pool
C. you step in it and get wet
D. it's a riddle
E. it's a puddle that rain makes....._____

4. GOWN
   A. you wear it
   B. something like an elf
   C. silk material
   D. an evening gown for a ball
   E. a lady's dress.....

5. EYELASH
   A. it's a hair on the eyelid
   B. it protects your eye
   C. an eye disease
   D. a horse whip
   E. you blink with it.....

6. MUZZLE
   A. black leather
   B. something to keep an animal from biting
   C. like a leash
   D. covering for an animal's mouth
   E. it's a fight—a quarrel.....

7. LECTURE
   A. it's a game
   B. a long composition
   C. when a man talks
   D. it's a speech
   E. a stage platform.....

8. SKILL
   A. performance
   B. something to cook with
   C. what you do very well
   D. acrobatic skill
   E. an ability.....

9. PECULIARITY
   A. a queer person
   B. when something odd occurs
   C. unusualness
   D. falsehood
   E. it happens very rarely.....

10. STAVE
   A. an oven
   B. it's like a staff
   C. you lean on it when walking
   D. it's wooden
   E. something you carry.....

Twenty-six per cent of the choices were characteristic of Feifel's higher conceptual levels or more abstract approach—synonyms and explanations, forty-six per cent were characteristic of the lower conceptual levels or concrete approach—use and description, and repetition, illustration, and inferior explanation, and twenty-eight per cent were errors.

Eight of the fifty choices were more 'difficult,' in the sense of being less frequent in occurrence in written English, than the stem word. In an ideal test none of the choices would have been less frequent in occurrence than the stem.

There were thirty-seven choices classified as use and description, repetition-illustration-or-inferior explanation, or error. The criterion set was that none of these choices characteristic of the lower conceptual levels be less 'difficult' than the synonym or explanation type response to the item. In the test there were fourteen of these thirty-seven choices which were less 'difficult,' i.e., more frequent in occurrence in written English, than the corresponding synonym or explanation type response. The other twenty-three choices characteristic of the lower conceptual levels were no less 'difficult,' i.e., no more frequent, than the synonym or explanation type responses.

PROCEDURE

The multiple-choice vocabulary test was given to a class at the third, fifth, seventh and eighth grade levels in a Brooklyn public school. Because the test requires that the pupils read the items, the lower end of the age range was necessarily limited.

The following directions were given to the class teachers who administered the tests to their own classes:

> Many people believe that children of different age groups define the same words in different ways. This test has been designed to study this theory. More than one answer is correct for each question on this test, but the children are to choose that answer which they consider best. It is to be emphasized that there is no 'right' answer.
>
> The teacher should read the directions with the younger groups. Stress the fact that the children should choose only one answer for each question. After reading the directions, allow five minutes to take the test. Collect all papers at the end of five minutes.

In addition to the public school sample the test was administered to a number of college graduates in order to compare their results with the responses of the children and with the responses typical of the higher conceptual levels, as found by Feifel and Lorge.

The total sample to which the test was administered is described in Table 2.

TABLE 2.—SAMPLE TESTED BY GRADE LEVEL AND MEDIAN AGE

Grade Level	Number	Median Age
3	37	8
5	38	10
7	29	12
8	30	13
College graduate	15	21+

The per cent of each type of response chosen by subjects at each grade or age level was obtained—based upon the total number of responses, including omits and errors. Because it was believed that this method of computing per cents might penalize the younger groups for whom the error and omit scores were rather high, the per cents were recomputed based upon the total number of 'correct' responses for each grade, where a response was considered correct if it would have been correct on the Stanford-Binet Form L vocabulary test.

The hypotheses to be tested, based upon the Feifel-Lorge findings (*2*) were that the per cent of synonym and explanation type responses would increase from grade to grade, that the per cent of use and description response would decrease from grade to grade, and that the per cent of repetition, illustration, and inferior explanation type responses would increase to age ten (grade five) and decrease thereafter. Significances of differences between per cents were determined by the critical ratio technique, using a one-tailed test.

RESULTS

The number and per cent of times each type of response was chosen by each grade are presented in Table 3. In the test itself, twenty per cent of all choices are synonyms. Grade three chooses synonyms as twenty per cent of all their responses. There is an increase in the per cent of synonyms chosen from grade three to grades five, seven, eight and college graduates, with a slight reversal between grades five and seven. Grade seven appears to be somewhat out of line with the other grades, suggesting that the students of this grade may not be equal in ability to the other students in the sample.

TABLE 3.—NUMBER AND PER CENT OF TIMES EACH RESPONSE TYPE WAS CHOSEN BY EACH GRADE

Grade	Synonym		Use-Desc.		Rep.-Ill.		Explanation		Error		Omit		Total	
	N	Per Cent	N	Per Cent	N	Per Cent	N	Per Cent	N	Per Cent	N	Per Cent	N	Per Cent
3	73	20	75	20	55	15	11	3	48	13	108	29	370	100
5	154	41	89	23	87	23	25	7	18	5	7	2	380	101
7	114	39	74	25	57	20	23	8	15	5	7	2	290	99
8	159	53	71	24	28	9	25	8	14	5	3	1	300	100
College	113	75	27	18	4	3	5	3	1	1	0	0	150	100
Test	10	20	14	28	9	18	3	6	14	28			50	100

The per cent of use and description responses remains about the same for all grades except the college graduates who show a slight decrease in choice of this type of response. None of the grades choose as many use and description type responses as are represented in the test itself.

The per cent of repetition-illustration-inferior explanation type responses increases from grade three to five, and decreases thereafter. The per cent of explanation type responses increases to grade seven, stays the same at grade eight, and drops for the college graduates. The per cent of error responses is never as high as the per cent of error choices in the test, and from grade five upward the per cent of errors never exceeds five per cent of the total number of responses. Finally, except for grade three, the number of omits is negligible.

The per cents of each type of response chosen by the different grades were compared by means of the critical ratio technique—the significance of the difference between two per cents. The results, summarized in Table 4, indicate:

1) The per cent of synonym type responses increases with grade or age, the difference becoming statistically significant between grade three and the higher grades, and between grades three through seven and the college graduates.

2) The per cents of repetition-illustration-inferior explanation responses differ significantly between grade five and college graduates indicating that such responses reach a peak at grade

five or age ten which corresponds to the peak found by Feifel and Lorge at age ten.

3) The per cents of use and description responses and of explanation responses do not differ significantly between any of the grade levels tested.

TABLE 4.—CRITICAL RATIOS OF PER CENT OF EACH RESPONSE TYPE CHOSEN BY EACH GRADE

	Grade			
	5	7	8	College
Grade 3				
Synonyms	2.10*	1.73*	3.00**	4.23**
Use and Description	.33	.50	.40	.17
Explanation	.80	.83	.83	.00
Rep.-Ill.-Inf. Exp.	.89	.56	.75	1.71*
Error	1.33	1.14	1.14	2.00*
Grade 5				
Synonyms		.17	1.00	2.43**
Use and Description		.18	.10	.42
Explanation		.17	.17	.67
Rep.-Ill.-Inf. Exp.		.30	1.56	2.50**
Error		.00	.00	1.00
Grade 7				
Synonyms			1.08	2.61**
Use and Description			.09	.54
Explanation			.00	.71
Rep.-Ill.-Inf. Exp.			1.22	1.89**
Error			.00	.80
Grade 8				
Synonyms				1.57
Use and Description				.46
Explanation				.71
Rep.-Ill.-Inf. Exp.				.86
Error				.80

* Significant at .05 level (t = 1.65 for one-tailed test).

** Significant at .01 level (t = 2.33 for one-tailed test).

Because the method used above of comparing the per cent of each type of response per total number of responses would tend to give an unfair advantage to the older children who knew more word meanings, the per cents were recomputed on the basis of the total number of 'correct' responses. The question was then raised, "Among the correct responses do the older children choose qualitatively different responses than the younger children?" Computing per cents on the basis of total number of correct responses are there significant differences in the per cents of synonyms, use and description responses, explanations, and repetition-illustration-inferior explanation responses for the different grades or age groups? These revised per cents are presented in Table 5.

TABLE 5.—NUMBER AND PER CENT* OF CORRECT RESPONSES OF EACH QUALITATIVE RESPONSE TYPE CHOSEN BY EACH GRADE

Grade	Total N	Correct		Synonyms		Use-Desc.		Rep. Ill.		Explanation	
		N	Per Cent	N	Per Cent*	N	Per Cent*	N	Per Cent*	N	Per Cent*
3	370	214	58	73	34	75	34	55	26	11	5
5	380	355	94	154	44	89	24	87	24	25	7
7	290	265	93	114	42	74	27	57	21	23	9
8	300	283	94	159	56	71	26	28	10	25	9
College	150	149	99	113	76	27	18	4	3	5	3
Test	50	36	72	10	28	14	39	9	25	3	8

* Per cents based upon total number of correct responses.

All grades, except the third, choose about the same per cent of correct responses. For grades five through college graduates the per cents correct range from ninety-three to ninety-nine, with grade seven again appearing out of line with the other grades.

The per cent of synonyms increases from thirty-four in grade three to seventy-six for college graduates. The per cent of use and description responses decreases after grade three, remains

relatively constant for grades five, seven and eight, and drops again for the college graduates. The per cent of explanations increases from grade three to seven and decreases after grade eight. The per cent of repetition-illustration-inferior explanation decreases from grade to grade.

The revised per cents of each type of response chosen by the different grades were compared by means of the critical ratio technique. These data are presented in Table 6.

TABLE 6.—CRITICAL RATIOS OF PER CENT OF CORRECT RESPONSES OF EACH QUALITATIVE RESPONSE TYPE CHOSEN BY EACH GRADE

	Grade			
	5	7	8	College
Grade 3				
Synonyms	.91	.67	1.83*	3.00**
Use and Description	1.00	.67	.73	1.23
Explanation	.40	.67	.67	.33
Rep.-Ill.-Inf. Exp.	.20	.50	1.78*	2.87**
Grade 5				
Synonyms		.17	1.00	2.29*
Use and Description		.27	.18	.50
Explanation		.29	.29	.67
Rep.-Ill.-Inf. Exp.		.30	1.56	2.62**
Grade 7				
Synonyms			1.77*	2.43**
Use and Description			.09	.70
Explanation			.00	.86
Rep.-Ill.-Inf. Exp.			1.22	2.00*
Grade 8				
Synonyms				1.43
Use and Description				.62
Explanation				.86
Rep.-Ill.-Inf. Exp.				1.00

* Significant at .05 level (t = 1.65 for one-tailed test).

** Significant at .01 level (t = 2.33 for one-tailed test).

In general the findings were:

1) There is an increase in the choice of synonyms as correct responses from grade three through college graduates, the difference in the per cents of synonyms being significant between grades three–seven and college graduates.

2) There is a significant decrease in the per cent of repetition-illustration-inferior explanation type responses between grades three–seven and college graduates.

3) There are no significant differences between the per cents of use and description type responses and the per cents of explanation type responses chosen by the students in the different grades.

It seems appropriate at this point to compare the results of Feifel and Lorge's study (*2*) on the qualitative differences in vocabulary responses of children as revealed in a recall-type test with the results of this study on the qualitative differences in vocabulary choices of children as revealed in a multiple-choice recognition type test. Responses on both tests were classified into five categories: synonyms, use and description, explanations, repetition-illustration- and inferior explanations, and error. Feifel and Lorge compared the mean number in each response category for each age group from six to fourteen. In the present study the per cent in each response category per total number of responses and per total number of correct responses was compared for children in grades three, five, seven, and eight and for college graduates.

The results of the multiple-choice, recognition type test agree with those of the recall type test for the synonym, use and description, and repetition-illustration-inferior explanation type responses when the analysis is similar for the two tests—based upon the total number of responses. When the analysis is based only upon the number of 'correct' responses, the results agree for the synonym and the use and description categories. The results from the recognition test do not always reach the level of statistical significance reached by the recall test, due most probably to the smaller sample size and the limited length and difficulty range of the multiple-choice test. They do, however, indicate that recognition vocabularies, just as recall vocabularies, differ in quality as well as in range from one age or grade level to the next. The fact that the trends are similar for the two types of

tests suggests that a longer revision of the multiple-choice test may yield as valuable information as the recall test with a considerable saving of time.

The most significant finding is the fact that even though a definition of a higher conceptual level is presented to the young child he tends to choose the response characteristic of the lower conceptual level—his own conceptual level. The younger child will tend to choose the use and description and the repetition-illustration-inferior explanation type definitions even when more abstract definitions in the form of synonyms are presented to him and are within his vocabulary range. He thus responds to the test in terms of his own conceptual level. Certainly further research along these lines will contribute to better understanding of the development of thought processes over the age span.

SUMMARY AND CONCLUSIONS

1) A ten item multiple-choice vocabulary test in which three or four choices were correct but of different qualitative levels was administered to a public school class at the third, fifth, seventh and eighth grade levels, and to a group of college graduates.

2) Definite trends, and in some instances significant differences, were found between the younger children and the adolescents and young adults for synonym, use and description, and repetition-illustration-inferior explanation type responses.

3) Among the 'correct' responses the younger children tend to choose more use and description and more repetition-illustration-inferior explanation type responses than the adolescents and young adults. The older groups tend to choose more synonyms than the younger groups.

4) There are, therefore, qualitative differences in the choices of the 'best' response to vocabulary items—the younger children tending to choose more concrete definitions and the older groups more abstract definitions.

5) These findings closely parallel those obtained by Feifel and Lorge (*2*) who studied qualitative differences in vocabulary responses to a recall type test.

6) The results of this study suggest that a multiple choice test of the type used in this study might be used to obtain better understanding of the mode of thinking and the level of intel-

lectual functioning of the child. More intensive research in this area is certainly warranted.

REFERENCES

1) Herman Feifel. "Qualitative differences in the vocabulary responses of normals and abnormals." *Genetic Psych. Monographs*, 1949, 39, 151–204.

2) Herman Feifel and Irving Lorge. "Qualitative differences in the vocabulary responses of children." *J. of Educ. Psych.*, 1950, 41, 1–18.

3) Edward L. Thorndike and Irving Lorge. *The Teachers Word Book of 30,000 Words.* New York: Bureau of Publications, Teachers College, Columbia University, 1944.

THE VALIDITY OF THE GOODENOUGH DRAW-A-MAN TEST IN GREECE[1]

I. TH. PAPAVASSILIOU

Athens, Greece

Since its introduction in 1926, The Goodenough Draw-A-Man Test has been used in a number of comparative studies of national and racial groups. However, as Goodenough has pointed out (*2*) simply because the test is free from verbal requirements does not mean that it is equally suitable for all groups. She cites studies by Dennis (*1*) and Havighurst (*4*) on American Indians to illustrate this point. These investigators, it may be recalled, noted a marked sex difference in favor of boys in tribes where art work is chiefly the responsibility of the males. An additional circumstance concerning the 'suitability' of the test, which is more pertinent to this paper, is that although Huang (*6*) noted similarities in the developmental sequences of both Chinese and American children, Hsiao found it necessary to restandardize the test for use in China (*5*).

In evaluating the usefulness of her test Goodenough has written as follows: "Repeated studies have shown that when used with children of reasonably similar cultural backgrounds who are equally motivated to do well, the test is serviceable as a crude measure of 'general intelligence,' although the moderate self-correlations and correlations with outside criteria make it clear that it cannot serve as a satisfactory substitute for individual tests of the Binet type." (*2*, p. 399)

The majority of the national and racial groups studied thus far have differed markedly in cultural background from that of the American standardization group. These groups have included Alaskans, Alorese, American Indians, Bengalese, Chinese, East Indians, Maya Indians, Mexicans, and Negroes in French East Africa. Because of the widely different cultural backgrounds of these groups, we cannot use Goodenough test results as a basis for any definite conclusions, except as to the relative 'suitability' of the device for each group. In effect, this

[1] The writer is indebted to Robert F. Biehler of the Institute of Child Welfare at the University of Minnesota for editorial assistance in preparing this report for publication.

is a measure of the similarity of the American 'culture' with that of the group being tested.

Curiously enough, investigators have overlooked those racial and national groups which are most similar to the United States in cultural background. We mean, of course, European countries. While Draw-A-Man procedures have been used in several of these countries, they have not been scored with Goodenough norms. It would seem worth while to use the standard Goodenough Draw-A-Man in these countries, in order to test the relative 'suitability' of the device. We might expect to be able to use such results as a basis for making inferences as to the degree of similarity of cultural background (as reflected in drawings), and, in the event of close similarity, as a basis for making rough comparisons of 'general intelligence.'

This study was undertaken in an effort to evaluate the suitability of the Goodenough Draw-A-Man Test for Greek children, by comparing Goodenough test results, scored against American norms, with the Terman-Sakellariou intelligence test, scored against Greek norms. The purpose of this procedure was to try to determine if the American Goodenough is suitable for use with Greek children, or whether some modification in scoring is necessary (as was the case with the Stanford-Binet). An earlier study (*9*) with eighty-six children of Greece (forty boys and forty-six girls) established a correlation of +.63 between the Goodenough and the Terman-Sakellariou scales. We have also discovered a general tendency for lower IQ scores on the Draw-A-Man test among both normal and subnormal children.

The present study is concerned with these findings in greater detail, on a new and larger sample. The children studied were one hundred forty-one boys and one hundred forty-nine girls in the first and second grades of public schools in Athens. The ages of the children varied from six to eleven years, but the majority (90 per cent) were between six and eight years. The socio-economic status of the subjects, as indicated by father's occupation, was as follows:

Professional and semi-professional	5 per cent
Government employees	25 per cent
Retail and minor business	26 per cent
Unskilled	25 per cent
Unemployed	4 per cent
Orphanage children	15 per cent

The preponderance of children in the lower classifications is reflected in the average IQ (Terman-Sakellariou) for the group, which was 93.5.

The Terman-Sakellariou scale (which is the Greek revision of the Stanford Binet) was given to the children individually. The drawings-of-a-man were obtained in class situations through the use of the standard Goodenough instructions and technique. The results of this procedure are presented in Table I.

TABLE I.—RELATIONSHIPS BETWEEN TERMAN-SAKELLARIOU AND GOODENOUGH IQ'S

	Terman-Sakellariou IQ		Goodenough IQ		Correlation Coefficient
	Mean	SD	Mean	SD	
Boys	94.1	14.6	89.3	17.4	+.74
Girls	93.0	15.1	87.2	18.6	+.68
Total	93.5	14.9	88.1	18.1	+.70

A more revealing summary of these results is afforded by a breakdown of the distribution of discrepancies between IQ's derived from the two scales:

4	(1.4%)	Goodenough IQ higher than Sakellariou IQ by 30 points
5	(1.7%)	Goodenough IQ higher than Sakellariou IQ by 20 points
33	(11.4%)	Goodenough IQ higher than Sakellariou IQ by 10 points
120	(41.4%)	No significant difference
69	(23.8%)	Goodenough IQ lower than Sakellariou IQ by 10 points
37	(12.8%)	Goodenough IQ lower than Sakellariou IQ by 20 points
17	(5.9%)	Goodenough IQ lower than Sakellariou IQ by 30 points
5	(1.7%)	Goodenough IQ lower than Sakellariou IQ by 40 points

As can be seen, the most striking trend is the tendency for the Goodenough score to be consistently below the Terman-Sakellariou score. This fact, together with the correlation of +.70, which is almost exactly the same as that found between the Goodenough and the Stanford-Binet by American investigators (*7,8,10*), seems to suggest that while the Draw-A-Man test is suitable as a ranking device, some modification may be necessary before it can be considered to possess its maximum effectiveness as an indicator of level of intelligence.

The greater proportion of cases in which the Goodenough IQ's fall below the Terman-Sakellariou IQ's almost certainly is in part due to emotional disturbances of the children (*3*). It has been our impression, although we have not given Rorschach or other personality tests, that we have in our group a number of disturbed children. Their teachers have reported special problems, such as speech handicaps, emotional instability, and sexual and other behavior problems. Such has also been our impression from the examination interviews with the Terman-Sakellariou scale.

It seems possible to trace the lower Draw-A-Man score to two other factors as well: differences in drawing between Greek and American children, and lack of art education among Greek children. The nature of the differences in drawing will be discussed in a subsequent paper. It is sufficient to say here that some of the points on the Goodenough scale do not seem to us to be correctly placed, developmentally speaking, at least for Greek children. We have also noted such marked sex differences that separate scales for boys and girls may be required. The lack of art education probably also has some influence on the score. In most of the schools, no more than an hour or two a week is devoted to art education; and the majority of the children had never drawn a man before. This lack of instruction and/or practice, might be expected to lower the score, and, at the same time, also might be a partial explanation for the greater variability of the Goodenough scores.

To summarize, we can say that the Goodenough Draw-A-Man Test, scored with the American norms, appears to be generally suitable for use with Greek children; however, even though this fact emphasizes the similarities between the two cultures, some modification of the scoring system seems necessary because of

differences in the drawings themselves, and because of a lack of art education in Greek schools.

REFERENCES

1) W. Dennis. "The performance of Hopi children on the Goodenough Draw-A-Man Test." *J. comp. Psychol.*, 1942, 34, 341–348.

2) F. L. Goodenough and D. B. Harris. "Studies in the psychology of children's drawings: II 1928–1949." *Psychol. Bull.*, 1950, 47, 369–433.

3) H. C. Gunzburg. "The significance of various aspects in drawings by educationally subnormal children." *J. Ment. Science.*, 1950, 96, 951–975.

4) R. J. Havighurst, Minna K. Gunther, and Inez E. Pratt. "Environment and the Draw-A-Man Test; the performance of Indian children." *J. abnorm. soc. Psychol.* 1946, 41, 50–63.

5) H. H. Hsiao. "Hsiao's revision of Goodenough's test of intelligence for children." *Monogr. Psychol. Educ. Nat. Cent. Univ.*, 1939, 4, No. 1 (*Psychol. Abstr.*, 1941, 15, No. 2836).

6) I. Huang. "The psychology of children's drawings." Shanghai: Commercial Press, 1938. (*Psychol. Abstr.*, 1939, 13, No. 3889).

7) Dorothea McCarthy. "A Study of the reliability of the Goodenough Test of Intelligence." *J. Psychol.*, 1944, 18, 201–216.

8) Edna W. McElwee. "The reliability of the Goodenough Intelligence test used with subnormal children fourteen years of age." *J. appl. Psychol.*, 1932, 16, 217–218.

9) I. Th. Papavassiliou. "Notes sur l' application du test du bonhomme en Grece." *Z. Kinderpsychiat.*, 1950.

10) J. H. Williams. "Validity and reliability of the Goodenough Intelligence Test." *Sch. &. Soc.*, 1935, 41, 653–656.

BOOK REVIEWS

CALVIN P. STONE, EDITOR. *Annual Review of Psychology.* Volume 3. Stanford, Calif.: Annual Reviews, Inc., 1952, pp. 462.

In this third volume of the *Annual Review of Psychology* there are two important new departures in comparison with previous issues. These consist of the inclusion of foreign reviewers and a better coverage of foreign contributions. The editors hope that these trends may increase in future volumes. The topics covered in Volume 3 are: child psychology, learning, vision, hearing, somesthesis and the chemical senses, individual differences, personality, social psychology and group processes, industrial psychology, comparative and physiological psychology, abnormalities of behavior, clinical methods: psychodiagnostics, clinical methods: psychotherapy, counseling: therapy and diagnoses, educational psychology, statistical theory and design, and motivation.

In general the reviews in this issue are well done when one considers the limitations of space together with the number of the publications covered. There are, however, certain shortcomings which will occur to the reader: (1) One expects reviews of this kind to include critical evaluations of the studies covered. Such evaluations are very uneven from review to review. Too often the treatment of successive studies is more like a brief abstract than a review. (2) The practice of giving only the author and journal with omission of the title in the bibliographies is decidedly unfortunate. However, the editors suggest that this defect may be remedied in future volumes. (3) One of the earmarks of a satisfactory review consists of a brief summary at the end pointing up the major trends discovered in the survey. Although a summarizing statement may be a difficult thing to write, no review can be entirely satisfactory to the reader without it. Only two of the seventeen reviews in this volume attempt a summary.

MILES A. TINKER
University of Minnesota

JOHN W. FRENCH. *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors.* Chicago: University of Chicago Press, 1951, pp 278. (paper) $4.00.

The results of sixty-nine factor analyses are summarized and cross-indexed to provide a reference book on ability factors.

This monograph contains a summary of each investigation, with a description of tests and subjects and rotated factor loadings. Then there is a tabulation by factors, indicating what tests are loaded with each. Finally, we have a test index, name index, and factor index. The whole is prefaced by thirteen pages of text. There are eight-hundred test descriptions, which in itself constitutes no minor piece of labor.

Having all this material assembled under one cover will benefit those concerned with factor analysis. The book will be of use to the general student if he does not take the material too seriously. This reviewer cannot escape the impression that French does worship factor analysis unduly, and that his work is peculiarly biased. Whenever his work touches on any unsettled issue, he appears to accept as final whatever position has been tentatively favored by Thurstone. French looks forward to the day when the test-constructor has a file of tests, each measuring one factor of the mind, which imputes to factors a status in nature rather than looking on them as a convenient way of establishing a frame of reference. As in many other reports based on the centroid method, no attention is given to unique factors. Another evidence of possible bias is found where it is implied (page 6) that a study which does not arrive at simple structure is "unsuccessful."

French limits his summary to studies using rotation to simple structure, and says with scarcely an apology, "To many groups of workers in this field it may appear unfortunate not to include any of the analyses of Spearman, T. L. Kelley, Brigham, Holzinger, Swineford, Hotelling, and Thomson, or of workers using their methods." It may indeed! The summaries might have been improved if all studies had been rerotated in terms of some common criterion. This would have allowed use of data from studies not aimed toward finding simple structure.

The author interprets tests in terms of ability when another psychologist might see the tests as influenced by set or motivation. Thus he prefers to see others' 'Plodding' and 'Carefulness' factors as abilities. French's report on the Age factor is surely open to criticism on this score. The factor is represented in several tests of an analysis by Harrell. Harrell's report but not this summary makes clear that the loadings reflected differences in attitudes of younger and older workers in the particular

cotton mill studies. That is to say, the tests do not measure anything inherent in Age. In general, French speaks of factors as inherent in tests, not as functions of test and sample of persons.

The introductory pages contain some theorizing which is of interest but needs further thought. Such a statement as "genetic factors are best measured by aptitude tests and experiential factors by achievement tests" is either circular or naïve.

This summary will find its greatest service as a compendium of research by the Thurstone school of factor analysts.

LEE J. CRONBACH

University of Illinois

G. MARIAN KINGET. *The Drawing-completion Test.* New York, Grune and Stratton, 1952, pp. 238. $6.75.

The drawing-completion test introduced in this book consists of a blank containing eight spaces on each of which is presented a set of small graphic elements that the subject is asked to incorporate into drawings. This form, originated by Wartegg, has the merit of presenting almost completely unstructured material from the standpoint of the subject, thus allowing for a large range of content. Yet differences in the stimuli, such as curved or straight elements, single or multiple elements, etc., afford reference points for interpretation and comparison of the drawings. The test is suitable for administration to groups as well as to individuals.

The test is scored on several variables in three major areas; stimulus-drawing relation, content, and mode of execution. The derived quantitative scores yield a profile in four bipolar dimensions; Emotion (outgoing or seclusive), Imagination (combining or creative), Intellect (practical or speculative), and Activity (dynamic or controlled). A scoring blank is presented, which permits the scoring of each of the eight drawings on every sign, using half steps on a scale of 0 to 3. The scores pertaining to each of the bipolar functions are then summed to obtain the final profile. After the profile is obtained, the diagnosis is individualized by consideration of the configuration of scores, absent criteria, order of execution of the drawings, etc., and by taking into consideration the age, sex, and occupation of the subject.

The book is divided into four major parts. Part one presents the "Origin and Development of the Drawing-completion Test," including the development of the scoring blank, the personality schema, and the methods used for validation of the test. Since experimental verification of the scoring criteria used for drawing tests is often lacking, it is particularly gratifying to note that the scoring for the Drawing-completion test was established on the basis of an investigation of three hundred eighty-three 'normal' subjects, divided about equally between the sexes, and ranging in age from eighteen to fifty years. A three-fold criterion was used, consisting of a questionnaire, a forced-choice test, and a rating scale, all designed expressly to measure the psychological functions represented in the personality schema. However, apparently the author has employed the questionable expedient of using the same data both for validation of the test and for subsequent changes and elaboration of the scoring system. Thus the diagnostic value of the final scoring variables as presented in this book cannot be considered to have been validated.

Part two of the book, entitled "The Diagnostic Mechanism," presents the method of administration, the basis for interpretation, and the scoring procedure for the test. This is the most extensive portion of the book. Each scoring category is defined and discussed in terms of the personality characteristics which it is claimed that it reveals or indicates. In these interpretations, the author appears to have allowed her enthusiasm to carry her far beyond the personality schema previously presented.

The final two parts of the book consist of discussions of case studies with reproductions of drawings and some filled-in scoring blanks. Specific scoring categories are also exemplified with illustrative drawings to aid in standardizing the evaluations.

The author is to be congratulated for having designed a projective test which can be scored on objective criteria and for having attempted to establish the scoring on the basis of experimental evidence. However, as a test manual, this book leaves much to be desired. At least three major omissions make it impossible to evaluate the usefulness of the test, or to take advantage of the underlying experimental data in utilizing the test. In the first place, almost no statistical data are presented; in fact, the only results presented from the experimental study are the percentages of agreement between the three-fold criterion

and the drawing test on three of the four functions measured. It appears from the text that the stated agreement was computed solely on the relative weight of the polar aspects of these functions; i.e., whether the person was more outgoing than seclusive, etc. However, it is impossible to ascertain the exact method by which the results were obtained, or their statistical significance. The only facts which are evident are that there was a greater amount of agreement among the tests for the females than for the males, and for both sexes the validity was lower for Emotion and Activity than for Intellect. Considering the enormous amount of work which has evidently gone into the investigation of relationships between personality and drawing characteristics, it is unfortunate that the author did not deem it practical to present a more complete summary of results in this manual, nor make it available in published form elsewhere.

The second omission is that the author has made no attempt to provide any norms for the profile scores. She excuses this omission on the ground that "free drawings, like all products of creative activity, do not permit establishment of rigorous norms." However, in the case studies remarks such as the following are made: "The degree of Emotion and Imagination exhibited here is exceptional for a male, though admissible at the age level of A."

It is evident that without some idea of what is 'normal' for a given age and sex it would be difficult to draw such conclusions, and hence norms are implied even though they are not presented.

Lastly, no data are presented regarding the reliability of the external criteria used in developing the qualitative interpretations of the various scoring categories. This is particularly important, since these interpretations were derived from study of clusters of items and even differential responses to single items in the criterion tests (page 24).

As a preliminary manual to aid in the attainment of standard administration and scoring procedures and some common basis of interpretation this book should be of service to those who may wish to do research on this test. However, it is hoped that other material in the form of norms and validity data will soon be made available to aid the clinician in his interpretation of the drawings and his evaluation of the usefulness of the test.

GOLDINE C. GLESER

Washington University Medical School, St. Louis, Mo.

RAYMOND B. CATTELL. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist.* New York: Harper & Brothers, 1952, pp. 462. $6.00.

The author states that the book has been written to meet three major requirements. "First, it sets out to meet the need of the general student in science to gain some idea of what factor analysis is about and to understand how it integrates with scientific methods and concepts generally." This comprises the first part of the book entitled, "Basic concepts in Factor Analysis." The reader seeking only orientation will obtain it by a careful reading of the 108 pages included here. "Second, it is intended as a textbook for statistics courses which deal with factor analysis for the first time, either as an appreciable part or as the whole of the semester course." The second part entitled "Specific Aims and Working Methods," combined with the first part, will meet this second requirement. "The third objective of this work is to supply a handbook for the research worker, the student, and the statistical clerk which will be a practical guide with respect to carrying out the processes most frequently in use." The third part of the book "General Principles and Problems," augments the straight-forward, preceding sections by treating variations in experimental design, unresolved procedural difficulties, alternative computational methods, etc.

Cattell has met these requirements which he has set for himself. The resulting text is highly readable, even entertaining. The multiplicity of diagrams, charts and examples plus the questions and exercises at the end of each chapter make the book teachable at the advanced undergraduate level and above. The author has kept the almost inevitable complex mathematical formulations and proofs to the barest minimum, though he furnishes pertinent references for the interested reader. The student new to the jargon of factor analysis, will make frequent use of the well-prepared glossary. The author, as has been his wont in other writing, occasionally stops the reader dead in his tracks with neologisms and relatively unusual words. This Cattellian quirk will probably stimulate most persons though it may irritate a few. The latter group are reminded that a little frustration can do no harm and it may facilitate learning!

With respect to the whole field of factor analysis, the book is neither comprehensive nor eclectic. It couldn't be and fulfill the stated requirements within the confines of a work of usual textbook size. The following instances are cited more as illustrations of the previous statement than as criticisms: A better case can be made for orthogonal, as contrasted with oblique, rotation than is done in the text. Cattell presents a strong argument for obliqueness and conjectures that only rarely will the best fit to psychological data be an orthogonal one. He treats only indirectly and summarily the orthogonalist's argument of the case of conceptualizing factors which are uncorrelated. He would seem to make computational ease the fundamental argument pro orthogonal factors. Many factor analysts, however, would insist on going beyond orthogonal factors to uncorrelated factor scores using partialling techniques to insure the latter. The author gives little attention to the method of principal components and the advantages which inhere in this mode of analysis for some purposes. In an early chapter the author includes a brief summary of various non-factorial statistical methods. However, he fails to mention canonical correlation and this method probably relates more closely to factor analysis than any of the others, yielding, as it were, a general factor common to both predictor and criterion measures.

The work has two unique features in terms of content which combine to make the book an indispensable addition to the professional library of everyone concerned in any way with factor analysis. First, throughout the book fundamental importance is attached to appropriate experimental designs. Cattell presents a model relating the various factor analysis designs to various experimental hypotheses. Also, he has devoted much of the third part of the book to a lucid explanation of how to incorporate computation, reliability, and validation checks in the design of various kinds of experiments employing factorial procedures. Second, he presents an exposition, with illustrative data, of his much discussed P and O analyses. He also proposes two new factor designs, S and T.

For the reader unfamiliar with these literal designations a word or two may clarify matters. R is used to designate the usual factor analysis of a matrix of variables comprising measures or scores. The factors yielded are, presumably, underlying and

more fundamental variables than the superficial ones used to obtain the matrix. Q is used to designate the transpose of R. It is a factor analysis of a matrix of correlations between persons. Putting this another way, the measures are made the 'cases,' and the cases are made the 'variables.' The resulting factors may be viewed as 'man-factor types.' P is like R except that it is done on a single person on whom many measures have been accumulated over a period of time. O is the transpose of P and yields, as it were 'personal occasion-factor types,' since occasions are correlated instead of the test scores. As to the S and T analyses, Cattell says: "These two logically possible but as yet unused designs correlate respectively two persons on one test on a series of occasions, and two occasions on one test on a series of persons." The former might be said to yield 'people-reaction-factor types,' the latter, 'social occasion-factor types.'

Cattell sets forth at some length the kinds of experimental situations for which the various factor analysis designs are appropriate.

The format and typography are excellent. *Errata* are seemingly few. The reviewer caught only one which can have serious consequences for the unwary reader. There are two miscalculations in Table 17, p. 163, which cause an accumulation of error in the table and lead to the reflection of signs on the wrong variables. An appendix is provided containing instructions for doing matrix multiplication by using electronic calculators, a feature useful to those fortunate researchers who can get access to such equipment.

Anyone interested in factor analysis should add this work to his professional library. Instructors of factor analysis whose students have sparse mathematical background will welcome this text. Cattell has made another major contribution to the psychological literature.

W. J. E. Crissy
Queens College, Fordham University